university of groningen

DRIVERS OF RETAIL VACANCY PRE-COVID19:
THE EFFECT OF BUILDING AND LOCATION FACTORS

Abstract

Last recent years, retail vacancy has substantially increased in the Netherlands. A frequently mentioned cause is the rapid rise of e-commerce. A jeopardy is that retail vacancy could eventually lead to the deterioration of buildings and areas accompanied by a reduction in livability. However, little is known about the effects of particular building- and location-characteristics on retail vacancy. This study investigates whether store agglomeration, population density, and building energy efficiency have an influence on retail vacancy. After executing multiple logistic regressions with and without the panel structure taken into account, the results show very few significant effects of building- and location-characteristics on retail vacancy.

Keywords: *retail vacancy, building-characteristics, location-characteristics, logistic regressions*

COLOFON

Document	Master Thesis
Title	Drivers of retail vacancy: the effect of building and location factors
Version	Final version
Author	Michelle Oskam
Student number	S3184161
RUG email	m.c.oskam.1@student.rug.nl
Private email	michelleoskam@hotmail.com
Institution	University of Groningen Faculty of Spatial Sciences Master of Science in Real Estate Studies Landleven 1, 9747 AD, Groningen
Supervisor	Em. Prof. dr. Ed Nozeman
Date	25-08-2021
Disclaimer	“Master theses are preliminary materials to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the author and do not indicate concurrence by the supervisor or research staff.”

TABLE OF CONTENTS

1. **INTRODUCTION……………………………………………………………………………4**
   1.1 Motivation……………………………………………………………………………4
   1.2 Literature review………………………………………………………………………6
   1.3 Research problem statement…………………………………………………………7
   1.4 Methodology and data…………………………………………………………………8
   1.5 Outline…………………………………………………………………………………9
2. **THEORY……………………………………………………………………………………9**
   2.1 Retail vacancy…………………………………………………………………………9
   2.2 Causes and underlying factors………………………………………………………10
   2.3 Building and locational factors………………………………………………………12
3. **DATA & METHOD………………………………………………………………………17**
   3.1 Data……………………………………………………………………………………17
   3.2 Reliability of the data…………………………………………………………………20
   3.3 Operationalization of variables………………………………………………………21
   3.4 Descriptive statistics…………………………………………………………………24
   3.5 Methodology…………………………………………………………………………26
   3.6 Model building………………………………………………………………………29
4. **RESULTS…………………………………………………………………………………31**
   4.1 Findings simple logit model…………………………………………………………31
   4.2 Findings panel logit model……………………………………………………………38
5. **CONCLUSION & DISCUSSION…………………………………………………………40**
   5.1 Conclusion……………………………………………………………………………40
   5.2 Discussion……………………………………………………………………………41
   5.3 Research recommendations…………………………………………………………42
   5.4 Critical reflection………………………………………………………………………44

REFERENCES……………………………………………………………………………45

APPENDICES……………………………………………………………………………48
Appendix A: Operationalization variables…………………………………………………48
Appendix B: Histograms store agglomeration……………………………………………50
Appendix C: Testing logistic regression assumptions ……………………………………50
Appendix D: Regression results ……………………………………………………………52

1. INTRODUCTION

This section serves as the introduction to this research in which the societal and scientific relevance becomes clear. In addition, the problem statement and the main research question with the subsequent sub-questions are described. Lastly, a short description of the utilized data and methodology is listed and the outline for the rest of the paper is stated.

1.1. Motivation

Currently, physical retail stores are under pressure due to the rise of e-commerce, new retail formats, and demographic changes. After some years of declining retail vacancy in the Netherlands, it increased again for the first time in 2019 since 2015. The vacancy of retail stores increased from 6.7% in January 2019 to 7.3% in January 2020. After several years of recovery, it is almost back to the peak level of 7.5% in 2015 as could be observed in Figure 1 below (Locatus, 2020).

Figure 1. *Percentage retail vacancy in number of physical stores over time (Locatus, 2020)*

The vacancy in square meters increased even more strongly, from 6.7% to 7.7% in 2019 (Locatus, 2020). Besides, vacancy as a share of the total floor space was highest in offices and shops in 2019. Analyzing this vacancy in retail stores, 2.6 million square meters were not in use (CBS, 2019a). A cause of the increase in vacancy is that in 2019 the number of stores decreased even further, 40% more than in 2018. In total 3% of the stores closed in 2019 (Locatus, 2020). Two other reasons are that the take-up of retail properties by the catering industry and the conversion of buildings into housing or offices are both stagnating (Trouw, 2020; Locatus, 2020). Until 2019, the centers of the largest Dutch cities have performed relatively well. At the beginning of 2019, the vacancy rate there was 5.8% which was below the national average. However, this cluster of shopping areas showed the largest increase in vacancy during 2019. With a vacancy rate of 7.3%, the vacancy rate in these centers was at the national average in 2020 (Locatus, 2020).

Although vacancy is generally the problem of the owner of a building, social effects play a crucial role in retail vacancy as well. Vacancy becomes a problem for society if negative effects occur for the environment, the rest of the city, or parts of it. Therefore, vacancy problems extend further than the property exploitation of an owner. When several vacant buildings are present in a shopping street, it could make the entire shopping area unattractive. Consumers stay away and entire parts of a shopping area fall into a negative spiral. The current vacancy then entails greater and more diverse risks. This includes the risk of deterioration of buildings and areas which will be accompanied by an increase in crime and reduced quality of life. Vacant stores are ideal places to engage in illegal activities, such as hemp plantations and money laundering (RPC, 2020). In addition, the loss of value of vacant real estate affects institutional investors such as insurers and pension funds. This could lead to a local domino effect with new vacancies and more loss of value of real estate as result (Vastgoedmarkt, 2018). It could even demolish the image of an area or city this way.

The movements towards an experience economy and the continuing penetration of online shopping are the main causes for the oversupply in Dutch shopping cities (Dynamis, 2019). Especially, the number of physical stores on B and C locations across the Netherlands continues to shrink after closings by retailers and the consolidation of national and international chains (Bouwinvest, 2020). This trend is directing to more uniform shopping destinations which makes it important to retain local retailers for experience shopping. Furthermore, Covid-19 has caused a decline in turnover for a lot of retailers and more consumers prefer online shopping. Before Covid-19 16% preferred online shopping, now more than twice as many consumers (35%) prefer this. Retailers expected that from 2020 until a year later the retail vacancy would increase from 7% to almost 20%, that 15% of the non-food stores will disappear and that these disappearances will further increase to 25- to 30% from 2020 until two years later in 2022 (Retailland, 2020; Vastgoedmarkt, 2020).

In addition, the population continues to grow annually by 0.5% in the upcoming years while the number of households will rise by 800,000 by 2040 (Bouwinvest, 2020). This is partly due to the ageing population. The number of young and old single-person households is rising and urbanization is continuing as well (CBRE, 2020). These demographic trends will have an impact on retail destinations; rezoning districts to make retail areas more compact, viable, and healthy. The retail landscape is shifting more towards services because of the changing needs of consumers.

So, it is clear that there is a problem that has an incredible impact on the retail market. Various causes of retail vacancy and the seriousness of it with its problems are revealed by different companies and newspapers. However, little is known whether other aspects like characteristics of the building itself and the location of the store in a shopping area also influence retail vacancy and if so to what extent. Therefore, this research focuses on the drivers of retail vacancy in the Netherlands.

1.2 Literature review

It is essential to review the literature on what has been addressed as the central concepts under review; retail vacancy and the drivers of it. This section provides a concise overview of the scientific literature on this subject to thereafter define the research problem which serves as the foundation of the derived conceptual model.

Research has already been conducted in this field of interest, all from different perspectives and definitions for the central concepts of this research. Firstly, it seems crucial to mention that it is challenging defining vacancy in the correct matter since real long-term vacancy is not present in times of transitioning a building to another sector for example. Thus, the exact demarcation when to call a property vacant seems undefined. Myers and Wyatt (2004) have dealt with this issue and described real long-term vacancy as follows:

> Buildings vacant or partly used for more than six or 12 months represent an unemployed resource. The period of vacancy seems arbitrary as the range of unoccupied buildings is so heterogeneous in relation to both form and function that definitional aspects become unnecessarily complex. It is the recognition of the concept of unemployment – wasted capacity – that is crucial (Myers and Wyatt, 2004, p. 286).

Clearly, vacancy needs a negative connotation to be defined as structural vacancy. The question remains what the drivers of structural vacancy could be.

E-commerce is a rapidly emerging trend and can be considered as an alternative to in-store shopping. In addition, it could make traditional retail less attractive (Weltevreden & Rietbergen, 2007; Zhang et al., 2016). Obviously, according to some studies, there exists a relationship between e-commerce and the demand for physical retail stores. A slowing growth rate of commercial property sales and the accelerating vacancy rate of commercial properties are closely related to the immense growth of e-commerce (Zhang et al., 2016). This relationship offers a clear view of e-commerce being a driver of retail vacancy. The extent to which people prefer shopping online instead of visiting a traditional physical shop is proven to differ among transport mode users and the perceived shopping attractiveness. For car users, the extent of car accessibility to the city center influences the propensity to purchase in physical shops. For non-car users, the higher the perceived shopping attractiveness the lower the probability they will shop online and substitute shopping at city center shops with online shopping (Weltevreden & Rietbergen, 2007).

Others have investigated the presence of an influence of vacancy rates in surrounding communities and of the overall national vacancy rate in the retail sector on local retail vacancy rates

(Benjamin et al., 2000). The role of distance in retail location choice is studied as well; empirical literature found that shops are willing to pay higher rents for locations on a short distance from an anchor store (Gould et al., 2005). This way they could benefit from the higher consumer flows it generates. According to Liu et al. (2016), retail mostly only occupies the ground floor in tall buildings. The consumers have to incur transportation costs to get to a higher floor, this tends to be prohibitive for stores to locate there. Besides, the continued growth in retail space and disproportionate growth in turnover cause a decrease in floorspace productivity (Evers et al., 2014). Turnover is not always the goal, ultimately it is about floor efficiency (profit per square meter). Furthermore, Teulings et al. (2018) have highlighted that pedestrian behavior of visitors leads to a negative distance decay in retail rents. Negative demand shocks lead to unprofitable locations for retail use at the edges.

Besides, a distinction could be made between types of retail stores that leave the premises. Based on the purchase motives of consumers a classification of shops could be generated: grocery shopping, recreational shopping, and targeted shopping. The vacancy in inner-city centers is mostly a consequence of the quitting of stores in the segment of recreational shopping. The share in the total vacancy of this has increased from 51% in 2004 to 65% in 2014. These inner-city centers deviate from the cities and the remaining of the Netherlands, wherein in 2014 the vacancy is largest in the targeted shopping segment, namely 50% and fun shopping 30% (Evers et al., 2014).

So, a substantial quantity is already stated by academic literature about for example e-commerce in relation to retail vacancy and location characteristics in relation to retail rents. However, there is little evidence of location characteristics such as store agglomeration in particular areas, environmental address density in a retail area, and building characteristics such as the energy label of retail stores in direct relation with retail vacancy. Therefore, the focus of this research paper is on this white spot in literature.

1.3 Research problem statement

Retail vacancy could be explained based on many studies related to for instance e-commerce, accessibility, retail rents, and economic- and demography changes. Currently, Covid-19 has a major impact on retail vacancy as well. However, there is insufficient knowledge in literature between the relation of building- and location-characteristics and retail vacancy. So, the research aim of this study is to offer insight into the importance of building- and location-characteristics in relation to retail vacancy. To arrive at the research objective, the following central research question will be investigated;

What are the drivers of retail vacancy in the Netherlands?

The three sub-questions that will be of support in answering the central research question are the following;

1. What drives retail vacancy and what is the effect of building and location factors?
2. What are the empirical effects of building- and location-characteristics on retail vacancy?
3. How do effects of building- and location-characteristics on retail vacancy differ between different residences with three different types of central shopping areas that are included for this research?

1.4 Methodology and data

The first sub-question will be investigated using academic literature. To answer this question, academic literature about the factors that influence retail vacancy will be reviewed and explained in detail. The second sub-question will be answered by conducting quantitative research. This will be achieved by running multiple logistic regression models. This method examines the influence of the independent variables (building- and location-characteristics) on the dependent variable (retail vacancy) as could be observed in the conceptual model in Figure 2. It quantifies the extent to which building- and location-characteristics are related to retail vacancy. Besides, control variables are included in the conceptual model (Figure 2). E-commerce, the economy, developments in population, and regional expenditures change the relationship between the dependent and independent variables. With the complete model, multiple regressions could be run. The last sub-question will be explored by testing whether there is a statistically significant difference between the different residences with the three different types of central shopping area groups.

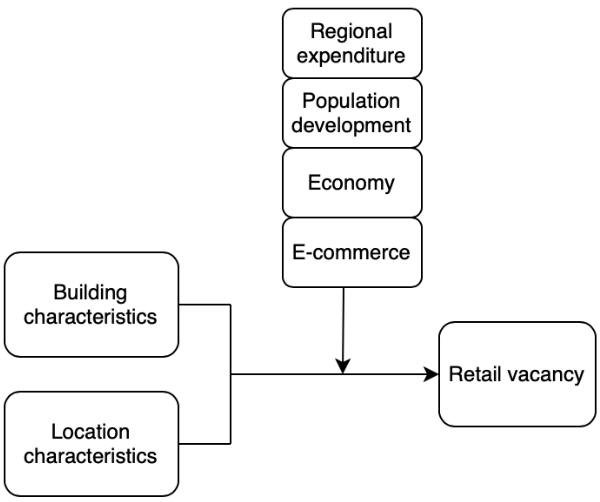

Figure 2. *Conceptual model (Author, 2021)*

For the central concepts of this empirical research secondary data will be used and retrieved from research organization Locatus and ESRI Nederland. Panel data are retrieved from 2012 till 2020 to research retail vacancy pre-COVID19. So that it is possible to check whether the building and location characteristics vary in this period. The dataset consists of addresses of the total retail stock in all residences with an inner-city or main shopping area in the Netherlands which includes both occupied and vacant stores with the floorspace, characteristics of the total retail stock, characteristics of the economy, urbanity level, and the age and quantity of inhabitants in the area in which the store is located. These characteristics of the total retail stock are for comparison to statistically prove the impact of building and location characteristics. Furthermore, secondary data from the CBS will be used as the control variables.

1.5 Outline

The remainder of this paper is organized as follows. Section 2 reviews and explains the existing literature regarding the main concepts in this research in which the first sub-question will be discussed. A theoretical framework will be outlined in which the causes and underlying factors of retail vacancy will be stated. This section will be finalized with the formulation of hypotheses based on literature. In section 3 the research design will be elucidated describing the different datasets to be used for the empirical research. Furthermore, the methods utilized and models built for this paper will be described as these are the foundation for the findings. Section 4 will present the findings of this empirical research in which the interpretation of the different models will be explained, the second and third sub-question will be answered here. Lastly, a conclusion will be drawn based on the theoretical and empirical findings, and afterwards in the discussion a critical reflection on the research will be stated with additional suggestions for follow-up research. This will be discussed in section 5.

2. THEORY

The theoretical framework is crucial for understanding the research topic and creating a knowledge base for the findings of the empirical part of this research. In this way, a thorough conclusion could be drawn based on a wider theoretical context. This section will firstly discuss retail vacancy, followed by several causes and underlying factors that influence retail vacancy. The section will be finalized by presenting hypotheses based on the discussed literature on building and location factors.

2.1 Retail vacancy

Retail vacancy is a visible indicator of how well a shopping street is doing. Often used indicators of the health of the retail industry as a whole are town center footfall, high street vacancy rates, and retail sales in specialized stores (Rhodes, 2014). Vacancy rates show large variations along primary and secondary locations in town centers. In prime shopping areas, vacancy rates are much lower than in secondary locations in the United Kingdom for example. Many retailers find themselves in the most difficult trading environment. A substantial number of retailers squeezed their margins, because of the inability to raise prices due to cheaper prices that are presented by grocers and especially e-commerce retailers. In the meantime, for some time retailers have to deal with increased rents, business rates, minimum wages, and in some cases the cost of raw materials. Besides, proportionately occupancy costs are highest for standard shop units which could explain the greater concentration of tenant administrations and store closures in this retail segment (Genecon et al., 2011). In 2011 in the UK there was a surplus of in-town secondary retail floorspace, of which a large quantity was no longer fit for purpose. Therefore, more quality shopping floorspace in high streets was needed, otherwise vacancy rates would increase even further (Genecon et al, 2011).

Another important characteristic of vacancy is its duration. In general, vacancy of less than a year is considered as frictional vacancy. Frictional vacancy is necessary for a market to function properly. It ensures that people and companies who are looking for space do not have to wait unnecessarily but are immediately offered the opportunity to meet their space needs. In the real estate market vacancy between one and three years is classified as long-term vacancy and from three years as structural vacancy (Evers et al, 2014; Locatus, 2018). In particular, structural vacancy is viewed as problematic and for this reason eligible for transformation or demolition (Evers et al., 2014).

However, shopping emerged in the first place as a major leisure activity in cities where customers could go out to explore new consumer spaces physically. This refers to the social aspect of shopping. The presence of diverse retail, cultural and other leisure facilities shape city centers as attractive destinations for recreational shopping (Weltevreden, 2006). This may offset the impact of e-commerce at these locations. So, e-commerce will not simply substitute city center physical retailing.

2.2 Causes and underlying factors

The supply and demand of retail space are critically important to understand for this research. Moreover, especially important for people and organizations owning, operating, and financing retail space. Supply and demand for retail space could be influenced by changes in retail sales, rental costs, land-use regulation, land availability, and the cost of capital (Benjamin et al., 1998). When supply increases and demand decreases; rents fall, and vacancies rise. Retail supply and demand are generally not in equilibrium given the reason of long lead times for retail space construction (Benjamin et al., 1998). Retail vacancy is the result of an imbalance in a market in which the stores that form the supply, respond inelastically and slowly to demand. Positive demand shifts will result in building new shopping space, however consequently this additional market supply because of the increase in demand will slow down the increase in rents and raise vacancy rates. Similarly, on the supply side, reasonable supply shifts decrease rents and raise vacancy rates as a result of a relative price inelasticity of retail space demand (Benjamin et al., 1998). Adding space to the total retail stock is easy to achieve, however extracting space from the total retail stock is almost impossible. Mostly, direct costs of the physical demolition or transformation are connected to this. So, both the decrease in demand and increase in supply has triggered vacancy rates to increase. It is the imbalance between supply and demand which affects the increase in vacancy in the retail market over the years.

Bringing to light the consumer demand during the Global Financial Crisis 2008-2014, retailers were hit by an extensive and lengthy negative consumption shock (McKibbin et al., 2009). This drop in consumption would result in a decrease in rents, a rise in vacancies, and transformations of land use from retail to other functions especially at the edges of shopping areas. To provide an indication, rents fell by 20% between 2008 and 2014 and vacancies rose by a factor of 1.6 in the Netherlands as a consequence of the macro-economic changes during that time (Teulings et al., 2018). The financial crisis has had an impact on the purchasing power of the Dutch population. This is of great importance when it comes to the development of retail vacancy in a shopping area as the purchasing power determines whether consumers still have room to be able to consume.

Currently, the rapidly emerging trend of e-commerce influences physical retailing. The increasing number of online shops and the huge growth of online sales threaten the performance and development of a great number of conventional physical stores. However, it is not likely that e-commerce will substitute in-store shopping and lead to the death of physical stores (Ward, 2001; Zhang et al., 2016). Especially in inner cities this could be stated, since people visit inner cities not only to go shopping, as well to meet and recreate. City centers account for approximately 52% of all retail outlets in the Netherlands, hereby these locations are the largest and most diversified shopping destinations (Weltevreden, 2006). The attractiveness of inner cities depends on several factors; a characteristic environment which is often historical, a concentration of a great variety of functions besides shopping, the number and variety of shops, and the crowdedness (Buursink, 1996). So, inner cities have an advantage over other shopping locations since customers visit inner cities for many different activities which are located in one central place.

As a consequence of the e-commerce trend, many retailers pursue an online store in addition to their physical store, which is called a ‘bricks and clicks’ strategy. Retailers pursue this strategy because of the interrelationship between the two and in order not to lose any sales and to keep competitive. Moreover, the responses to e-commerce differ between individual retailers and consumers, owing to variation in individual preferences, available resources, and the context in which an actor operates. As with any change in the retail landscape, improvements in technology are not the only driver. Other trends and developments in the environment, demography, social, lifestyle, economy, and politics all influence the future prospect and use of e-commerce. E-commerce could not be separated from the wider contextual movements shaping the retail sector (Burt and Sparks, 2003).

The continuous ageing of the Dutch population, a growing need for fun shopping, and the growth of out-of-town retailing are as much of importance for the future of the retail sector. The ageing of the population leads to decreased consumer spending (Hurd and Rohwedder, 2003; NRW, 2014). So, the increase in market size for retailers is no longer an effect of population growth. Although for decades single-person households rise and take place particularly within the stagnant population in peripheral regions, consequently this has led to a higher need for household goods such as washing machines and kitchen amenities. This increase in one-person households is forecasted to keep rising in the near future (CBS, 2019b), which will presumably lead to an even higher demand for household goods.

2.3 Building and locational factors

So far, an attempt has been made to present the appropriateness of supply, demand, e-commerce, demographics, and other underlying retail factors as a theoretical base for examining the impact of building and locational factors on retail vacancy. Hypotheses will be formulated concerning the relation between building- and locational factors and retail vacancy, based on literature.

The location of a physical retail store is important for the relation with its market and as a competitor to other retailers. Geographically, real estate markets are highly segmented (Geltner et al., 2001). From the perspective of a retailer, the clustering of dissimilar stores shows more than doubling profits compared to stores that are at independent locations (Ghosh, 1986). In these locations with the agglomeration of different stores, consumers benefit from multipurpose shopping since it is cost reducing in a timesaving and in an energy-effective way (Kumar and Karande, 2000). However, Caroll (1985) proposes that although specialistic organizations do not compete with general merchandise retailers for secondary resources, they could have difficulties obtaining the central resources. Thus, resource partitioning suggests that specialistic retailers do not benefit by locating close to general organizations. The agglomeration of stores selling different but complementing products could benefit from clustering though. According to Brouwer and Tool (2018), more diversity of the range of shops in the main urban shopping area is linked to less retail vacancy. So, contradictory statements on store agglomeration that apply to differences in retail mix are brought into light by different studies in empirical literature.

The desirability and choice for a store to a consumer depend on the attractiveness of a specific store regarding merchandise quality, service quality, ambiance, and deal proneness. Customer satisfaction which results from these determinants is likely to generate consumer loyalty, which is important to achieve for retailers (Paul et al., 2016). However, these are not the only factors consumers consider, the location and distance of the store from the customer and the relative distance to other stores in the neighborhood are considered as well. This suggests that distances from competitors of the same type are important to consumers. Retailers keep in mind such information in determining a location strategy. Minimum differentiation of stores in an area results in uncertainty-, time-, and search cost reduction and facilitates comparison-shopping, which could benefit retailers as well (Kumar and Karande, 2000; Karande and Lombard, 2005). Consumers could be attracted to a wide variety of the same product category which may differ in price, size, and design in multiple comparable stores. The clustering of retail activities ensures advantages for retailers that include sharing infrastructure and the benefits of the stream of consumers (Teller and Elms, 2010). Competition amongst retail facilities has increased and retailers often tend to locate close to competing stores. Furthermore, the proximity of retail stores has to do with the following factors: proximity is higher in areas with high income, high population density, greater retail expenditures, younger population, and high homeownership. On the other hand, retailers tend to distance themselves from each other in areas with the opposite market characteristics. Moreover, when retailers experience high overlap of merchandise, customers, labor, and financial needs among competing retailers then a distancing strategy is preferred (Karande and Lombard, 2005). Karande and Lombard (2005) analyzed these proximity and distancing strategies for retailers in the following two categories; home improvement and office products, since these retailers are present in many geographical markets. In their research, the focus was on three similar-midsized markets in the USA. These markets are intensely competitive and the advantage of stores offering similar products is greatly influenced by location. Besides, general merchandise stores were taken into account in this research. Using statistical analyses, the aforementioned strategies were examined. However, the direct link with retail vacancy is missing. The clustering of different stores could benefit customers and retailers in several ways, but it could also harm retailers. So, based on the foregoing evidence, the first hypothesis follows;

H1: *The more stores locate close to competing retailers in an area, the less likely it is stores becoming vacant.*

PBL and ASRE (2013) have empirically determined that retail vacancy and as a result deterioration, mainly occurs in so-called secondary and tertiary shopping areas in inner cities, such as entrance and ring streets. The main reasons for this are consumers cannot go fun-shopping and recreate in these areas, retailers want bigger stores and the floorspace of the retail stores in these areas do not meet these requirements, and both municipalities and project developers prefer to develop a new area rather than to renovate an existing one. Thus, retailers recognize the store location decision is of great importance for their long-term success. After deciding about the size, design, and operational requirements of a site, the retailer identifies key location criteria. A study by Schmidt (1983) at the University of Colorado at Denver identified several main site characteristics that were desired by retailers. These characteristics include high traffic volume, maximum street frontage, wide curb cuts, safe access to traffic in both directions, parcel size, and community population threshold. Brubaker (2004) identified site selection criteria for retail tenants. These included signage, visibility, traffic counts, parcel size, parking, co-tenancy, proximity to other attractions like restaurants and theaters, demographics such as population, income, education, and finally competition and trade area. More recent empirical literature highlights criteria for selecting a store location in seven categories; performance measures, population structure, economic factors, competition, saturation level, magnet and store characteristics (Turhan et al., 2013). So, several studies highlighted criteria for choosing a store location, this has been done by conducting interviews with specialists or performing a literature review. Overall, the characteristics that seem crucial in most of these studies are population, store size, accessibility, and traffic volume. However, retail site selection remains a decision process of feeling and gut feeling based on experience. Always the risk remains of not picking the right site. Although, nowadays there are advantageous factors for the success of site selection such as technology, knowledge about retailers' sales and trade areas estimates. When analyzing population density and urbanity, research organization Locatus focuses on the proximity of people in an area. It demonstrates this with the number of inhabitants within a radius of two, five, and ten kilometers. When the number of inhabitants within the radius of for example two kilometers becomes higher, the liveliness of the area is higher and can be experienced as pleasant, cozy, and attractive (Wesselink, 2004). In these areas, the building density, the proximity to amenities, competitors, public transport all tend to be greater which is usually experienced as pleasant. Population density has a positive impact on liveliness and that in turn is experienced as pleasant and attractive. However, it does not state anything about the performance of a retail area directly, so no conclusions could be drawn about retail vacancy. Furthermore, as already stated the crucial characteristics for a store location are determined by conducting interviews. So, this research will give one of the first insights of statistical evidence on the influence of the specific location factors on retail vacancy. When encompassing the crucial characteristics from literature as population, accessibility, and traffic volume the second hypothesis is as follows;

H2: *The higher the population density in the catchment area, the less likely it is stores becoming vacant.*

To further elaborate on the accessibility, empirical literature investigated the effect of the time it takes to travel from home or work to a shopping area and the number of visitors in a shopping area. The lower the time spent on this, the higher the number of visitors. Furthermore, the type of transport and the number of obstacles encountered along the way are important factors that could influence the choice of the consumer (Teller and Elms, 2010). Obstacles include, for example, delays due to traffic jams and traffic disruption. A special type of obstacle that is important for car users is the possibility of parking. The availability of free parking spaces and the type of parking facilities offered is considered as an integrative part of perceived accessibility of a retail destination. When parking facilities do not correspond to the size of the shopping area so that there is shortage of parking space, it becomes difficult to reach the shopping area, which increases the risk of vacancy. So, the purchasing motive of convenience is by no means always facilitated, this thus includes accessibility, parking, compactness (NRW, 2014). Among retailers there exists a belief that parking plays a fundamental role in the performance of a shopping area, consumers regard a visit to a certain shopping area as a deterrent if there is reduced parking capacity and/or increased parking tariffs. However, Mingardo and van Meerkerk (2012) found that higher parking fees are associated with higher turnovers per sales floor surface in shopping areas, which is in contrast to what retailers generally believed. Although, according to Mingardo and van Meerkerk (2012) there is no statistical relationship between parking capacity and turnover except for regional shopping areas. These findings on the price and quantity of parking are based on cross-sectional data analysis with log-linear regression analysis. However, the perceived safety or atmosphere of a shopping area that may have a strong underlying role is not included as a control variable.

Parker (1992) distinguished dimensions of the physical attractiveness of stores internally and externally; the range of goods in a store, the service component of the offer, the relative accessibility of the store to the target consumers, and consumer attitudes towards the environment of the city center in general as a shopping destination are considered as significant. Store-image attributes explain a sizeable amount of the store performance, these include atmospherics, assortments, quantity and quality of products (Turhan et al., 2013). Retailers should look into store characteristics to gain competitive advantages or better performance against their competitors in the market. Improving these characteristics has an impact on both revenue flows and expenses. So, it remains unclear whether improving the store-image attributes result in higher store profits and thus in lower retail vacancy.

Besides, the average floorspace per store has increased significantly in recent decades, from less than 50 square meters per store in 1968 to more than 270 square meters in 2013. This is largely due to the chain formation that has taken place in the retail market. This need for large-scale stores is also reflected in the vacancy figures. Vacant retail buildings are on average smaller than buildings in use (PBL & ASRE, 2013). Larger units or anchor stores have a strong position in the retail market. These large parties are also referred to as magnets since they have the capacity to attract visitors and smaller retailers (Damian et al., 2011). Due to this increase in scale, attention is mainly given to the establishment of large stores while the attention for small-scale buildings is disappearing. This means that there is a substantial quantity of vacancies in this type of retail properties.

Looking at store performance from another angle, the environmental impact of a building stock could be analyzed. The commercial property sector must as well adhere to governmental regulations concerning the energy performance of a building. Part of this policy contains compulsory energy labelling of commercial properties at a transaction moment (RVO, 2018). Furthermore, this sector has seen the emergence of fiscal incentives, voluntary business responses, and industry standards. This is mainly happening in response to environmental pressures to reduce the environmental impact of the total building stock. Benefits from building energy efficiency for tenants mainly include lower utility bills and other financial incentives. More intangible benefits may include improvements in business and marketing performance. However, Fuerst and McAllister (2011) highlighted that there is no evidence that energy labelling had any effect on market rent or market value of commercial property assets. There are no significant effects found on market rent or market value associated with different energy label ratings. Their study does confirm that properties tend to decrease in value as they get older. Important to acknowledge is that the sample size was relatively small, the assets included all sectors of the commercial market and were spread across all commercial property regions of the UK. Thus, statistically significant differences between different sub-groups are not likely to be observed. To test the expectation that energy efficient buildings lead to less retail vacancy, the third hypothesis is formulated;

H3: E*nergy efficient retail buildings are less likely of becoming vacant.*

Lastly, to summarize this section, an overview of the variables that directly or indirectly influence retail vacancy is listed in Table 1 below.

Table 1. *Overview variables influencing retail vacancy*

Demand side variable	**Supply side variable**	**Author + publication date**
Economy		Teulings et al, 2018
	E-commerce	Ward, 2001; Burt and Sparks, 2003; Zhang et al, 2016
	Urbanization	Buursink, 1996; Weltevreden, 2006
Demography		NRW, 2014; CBS, 2019
	Retail floorspace	Genecon et al, 2011; (PBL & ASRE, 2013)
	Store agglomeration	Ghosh, 1986; Kumar and Karande; 2000; Karande and Lombard, 2005; Teller and Elms, 2010
Population density		Wesselink, 2004
	Accessibility	Teller and Elms, 2010; NRW 2014
	Parking facilities	Mingardo and van Meerkerk, 2012; NRW 2014
	Store size	Damian et al, 2011; PBL & ASRE, 2013
	Building energy efficiency	Fuerst et al, 2011

3. DATA & METHOD

This section describes the set-up for the empirical study; the data used for the research with the corresponding sources, the reliability of the data, and the operationalization of the variables will be discussed. Furthermore, an overview of the descriptive statistics will be presented with all variables used in the regressions. Thereafter, the methodology will be explained and how it will be implemented in this research. Lastly, the model building process for this research is presented and explained.

3.1 Data

The data required to perform multiple regressions for testing the proposed research hypotheses are data on the dependent variable as well as data on the independent variables and control variables. The data come from different datasets and sources. The data from research organization Locatus are used for the central concept retail vacancy in this empirical research as the dependent variable. The research organization Locatus registers a retail property as vacant if: “It is reasonably expected that a sales point in the retail trade, catering industry or consumer-oriented service will return to the (vacant) building” (Locatus, 2018). In addition, there are other criteria. A building is only registered as vacant for buildings inside a shopping area if; the building was in use as a shop and is now actually empty, or the building is no longer in use as a shop or catering industry at that time, but it is indicated on the building that it is for sale/rent (as a sales point). Outside shopping areas, both criteria must apply; there must have been a sales point and the property must be for sale/rent or sold/rented out. The length of the vacancy period is not considered in this research since unavailability of this information in the data for all years. The vacant retail properties are retrieved from Locatus. This dataset consists of all vacant retail properties within a residence with inner-city or main shopping area in the Netherlands, including all different main types of shopping areas that are defined by Locatus in these selected residences. Several different central shopping areas are distinguished by Locatus. The first main type that Locatus defines is a central shopping area, this main type is divided into three different subtypes: inner city, a large- and small main shopping area. An inner-city contains more than 400 shops. Inner cities are the top 17 shopping areas in the Netherlands, including the inner cities of Amsterdam, Rotterdam, The Hague, Utrecht, Groningen, and Maastricht. Residences with a main shopping area include 121 residences. A main shopping center is the largest shopping area in the residence and is divided into two categories by Locatus. The first category is a large main shopping area which consists of 200-400 stores. Examples are the center of Bussum and Delft. The second category is a small main shopping area which consists of 100-200 shops. Examples are the center of Schagen and Putten (Locatus, 2017). The share of the number of retail outlets in these central shopping areas relative to the total retail stock in the Netherlands is 44.5% in 2020 which is a considerable high amount (Locatus, 2020). In total there exist 219,156 retail outlets in the Netherlands in 2020, of which 97,524 retail outlets are located in a central shopping area. The share of these retail outlets is the highest in the dataset used for this research. The other main types of shopping areas defined by Locatus such as supporting shopping areas, scattered shopping, and the category other, are only included in the dataset if these are located in the same residence as the ones with an inner-city or main shopping area. In 2020, the share of scattered shopping in the Netherlands is 34.1%, the share of supporting shopping areas is 18.7% and other areas cover a share of 2.6% (Locatus, 2020).

In addition to the vacant retail properties, occupied stores are retrieved from ESRI Nederland (2020) which is extracted from the 'Basisregistratie Adressen en Gebouwen' (BAG) in which all buildings in the Netherlands are registered with associated rights. The Geodienst, which is the spatial expertise center of the University of Groningen, eventually compiled the dataset with the occupied stores. Unnecessary observations are dropped on postcode4 level. These unnecessary observations include the occupied stores that fall outside the focus areas of residences with an inner-city or main shopping area. The Geodienst retrieved the energy labels as well for the occupied stores. The source of these energy labels is ESRI Nederland (2020) who has already prepared the labels of 'Rijksdienst voor Ondernemend Nederland' (RVO). Besides, the Geodienst compiled another dataset with all the occupied stores over the years 2012-2020 for which no current address is recognized. For some of these occupied stores an energy label is acquainted. These occupied stores are included to realize a more complete retail stock dataset since these are properties that had a retail function between 2012 and 2020 and currently no address is recognized anymore for these properties. This could occur during splitting or merging of retail properties, demolition or new construction, or change of function. Unfortunately, it appears to be impossible to link historical addresses. After combining all different data into one dataset some duplicates appeared which have been deleted from the dataset. An ID variable on property level has been created since initially there was no ID available for every observation on property level. If there are multiple stores in the same building that are registered at the same address, they have been removed from the dataset. These different stores could not be distinguished from each other when creating property IDs based on the street, house number, and city. No unique characteristic is known for these stores, so these appeared as duplicate observations in the dataset which thereafter have been removed. After cleaning all the data, the remaining total properties over time for the analysis concern 779.723 observations. For a simple logit analysis, the observations after the first detection of vacancy in a particular building are deleted from the dataset. In the dataset on municipality level the remaining data for analysis concern 706.794 observations and on postcode4 level it concerns 707.211 observations.

The independent variables include store agglomeration, population density, and building energy efficiency. Firstly, store agglomeration will be measured as the number of stores in a particular shopping area. This variable is created in Stata by counting the retail properties in a particular area. This is performed on two different levels: on postcode4 level and postcode6 level. Secondly, population density will be measured as environmental address density that indicates the number of residential addresses per km^2 (categorized in urbanity levels) in each catchment area on postcode4 level which is retrieved from CBS (2020). Lastly, building energy efficiency will be measured with the energy labels of retail premises retrieved from ESRI Nederland (2020) and Rijksoverheid (2020). The data from the Rijksoverheid could be observed and retrieved via the online Energy Performance Platform (EP-Online).

Data from CBS have been used for most of the control variables that have been taken into consideration to possibly include in the empirical model. The control variables include variables on economy, e-commerce, demography, ownership, income, and retail floorspace. However, some control variables possess too many missing values, causing these cannot be included in the analysis. Other variables were only available on macrolevel instead of microlevel, these could not control these effects and hence these are as well not included in the analysis, for example the e-commerce variable. Nevertheless, an indication will be given of all important variables that were (partly) available at CBS. Economy factors were available as consumer confidence, economic climate, and willingness to buy in the Netherlands; this includes the expectations of Dutch consumers, which is corrected for general economic developments and their financial situation. E-commerce is measured as the turnover development in internet sales in the retail sector in the Netherlands. Demography is measured as the key figures of the Dutch population indicating gender, age, households, and population growth. Ownership is measured as the distribution in percentage of rent and owner-occupied homes. Income is measured as the percentage of high and low household incomes. Lastly, retail floorspace for the vacant properties is retrieved from research organization Locatus, which is measured as “the square meters of a (retail) unit that are freely accessible or visible to the customer, including areas directly associated with the sale” (Locatus, 2019). For the remaining occupied retail stock, the square meters of the accommodation objects are adopted and retrieved from ESRI Nederland 2020.

To collect the data from research organization Locatus, the Geodienst of the University of Groningen, and CBS, Excel- and csv files are retrieved. Thereafter, these data files are imported and merged in Stata to be able to run the logistic regressions. All the data used for this research are panel data, which are retrieved from 2012-2020 to make a thorough analysis over time to check whether the building and location characteristics vary over this time. A clear overview of all variables with the corresponding indicators and sources that seem important for the analysis are shown in Table 2 below.

Table 2. Overview variable sources with indicator(s)and spatial scale

Variable type	Variable	Indicator(s)	Spatial scale	Source
Dependent	Retail vacancy	Vacant retail premises in all residences with a city center or main shopping area, indicated by address	Unit level	Locatus, 2020
Dependent	Total retail stock	All accommodation objects with retail as the function, indicated by address or postcode	Unit level	Geodienst, University of Groningen; retrieved from ESRI Nederland 2020
Independent	Store agglomeration	Number of stores on two area levels; in the same postcode4 and postcode6 area.	Postcode4, postcode6 level	Locatus, 2020; ESRI Nederland 2020
Independent	Population density	Number of environmental address density on postcode4 level, indicated as the total number of addresses per km^2. The urbanity level on postcode4 level and municipality level, indicated with category 1-5	Postcode4, municipality level	CBS, 2020a
Independent	Building energy efficiency	Energy labels of retail premises, with date of registration or provisional	Unit level	Geodienst, University of Groningen; retreieved from ESRI Nederland 2020; Rijksoverheid EP-Online, 2020
Control	Economy	Consumer confidence, economic climate and willingness to buy; the expectations of Dutch consumers, that is corrected for general economic developments and their own financial situation	National level	CBS, 2020b
Control	E-commerce	Turnover development in internet sales in the retail sector excluding petrol stations and pharmacies	National level	CBS, 2020c
Control	Demography	Key figures of the Dutch population on postcode4 level; gender, age groups. On municipality level; population density, population growth, and households	Postcode4 level	CBS, 2020a CBS, 2020d
Control	Ownership	Percentage of owner-occupied homes and of rental homes	Postcode4 level	CBS, 2020a
Control	Income	Percentage of high and low household incomes	Postcode4 level	CBS, 2020a
Control	Retail floorspace	Square meters of a (retail) unit that are freely accessible or visible to the customer, including areas directly associated with the sale. Square meters of the accommodation object	Unit level	Locatus, 2020 Geodienst, University of Groningen; ESRI Nederland 2020

3.2 Reliability of the data

Locatus has a standardized method of collecting their retail data, whereby all retail stores in all shopping areas in the Netherlands are visited periodically on a yearly basis. Not all shopping areas are visited at the same time. Thus, this results in a delay. In terms of vacancy duration, Locatus is dependent on the moment the field staff visits a building. A building could therefore already be vacant for 11 months at the time when Locatus visits the property. This will be registered as initial vacancy for another year, while it actually quickly becomes long-term vacancy. Furthermore, the demarcation of shopping areas entails a certain degree of subjectivity and inconsistency. Sometimes adjacent shops and shopping streets are included in city centers, while occasionally these are not included. As long as the parties adopt the same classification there is little to worry about, but otherwise it could affect the values of certain variables. For example, the number of stores in an area, the degree of compactness, and the vacancy rate. Despite these caveats, these are large datasets, which in terms of quality and reliability are the best of what is available in the market. Broadly speaking, it is allowed to assume that the data are sufficiently reliable.

Other caveats to deal with are that the research is limited by the availability of data on individual variables and unavailability of data on retail unit level, which are of considerable importance in predicting differences in retail vacancy based on scientific literature. For some control variables as urbanity, homeownership, and income, not all data are available on postcode4 level for every year in this analysis. To take these caveats into account; the analysis will be done twice, on postcode4 level with the missing data left out of the analysis and another time on municipality level for which more data are available for almost every year. Besides, for the e-commerce variable not every year is available, only the years 2014-2019 are available. If there are as well too many missing values on municipality level the variable should be deleted to lower the missing values to the minimum. The deleted variables are the following ones: homeownership, income, and e-commerce.

Furthermore, not for all properties with a retail function an energy label is recognized. In the final dataset in which the total retail stock is listed from different sources, approximately 38% of the energy labels are known. This may have negative effects on the representativeness of the outcome. Furthermore, there is a distinction between energy labels that are registered and not registered, these unregistered labels are labeled as provisional. A provisional label is an estimate based on average values, housing type, and year of construction. For each construction year and housing type, the RVO has established the most common energetic situation based on ‘woON2006’ (Berben and Kuijpers-van Gaalen, 2014). The provisional energy label is therefore not based on the actual building characteristic values of the property in question but contains an estimate of it that is as accurate as possible.

3.3 Operationalization of variables

Retail vacancy

The data provided by research organization Locatus, shows retail vacancy only in the three categories frictional, long-term, and structural vacancy, for the last three years. Before that time, it was only measured as yes, a building is vacant or no, a building is not vacant. Therefore, to make a thorough analysis over time from 2012-2020, the variable is measured as a binary variable. Here, the dependent variable retail vacancy has either the value 0 when the retail property is occupied and the value 1 if the retail property is vacant. In Table 3 in Appendix A the proportion with the percentage of the vacant retail properties could be observed in relation to the total retail stock for all years in total. In Table 4, which could be observed in Appendix A, an overview of the proportion of vacant properties related to the total retail stock is created in more detail per year. As could be observed in Table 4 the percentage of vacant properties increases every year. Overall, this is in line with the actual retail vacancy in the Netherlands, from 2012-2020 retail vacancy has increased. However, in the years 2015-2018, the actual retail vacancy slightly decreased. Thereafter, in 2019 it increased again. Besides, the percentages in the dataset show higher results than the actual percentages of retail vacancy in the Netherlands. This is because not every single retail property is included in the dataset, only retail properties that are located within a residence with an inner-city or main shopping area. This limits the validity of the analysis. In addition, the location of the vacant retail properties in particular shopping area types is captured. In Table 5 in Appendix A, an overview is given about the proportion of vacant stores in the different shopping area types. Logically, most vacant stores could be found in central locations since here most retail properties are located.

Environmental address density & urbanization

CBS uses the environmental address density to determine the degree of concentration of human activity: living, shopping, and working. The environmental address density is calculated using all addresses in the Netherlands. The environmental address density of an address is the number of addresses within a circle with a radius of one kilometer around that address, divided by the square meters of the circle. The environmental address density is expressed in addresses per square kilometer. The value of the environmental address density is calculated per square of 500 by 500 meters and assigned to all addresses located in this square (Leeuwen, 2019). As of 2015, the environmental address density is derived from the number of residential objects, berths, and stands in use from the BAG. Before 2015, addresses and their coordinates of the grid squares were derived from the Geographic Base Register (GBR). This register contains all addresses in the Netherlands, provided with the postcode, the municipality code, the district- and neighborhood code, and the coordinate of the 500-meter grid square.

Urbanity is a categorization of the environmental address density. There are five urbanity classes. The classes with the corresponding environmental address density and the distribution of it in the dataset could be observed in Table 6 in Appendix A. Only for the years 2014-2019 the environmental address density with the corresponding urbanity level on postcode4 level is available. The years 2012, 2013, and 2020 are missing. Therefore, this variable has only 505,055 observations. On municipality level, all years are available and then show 749,615 observations.

Building energy efficiency

Building energy efficiency is measured by registered and provisional energy labels. Only for the registered energy labels an energy index is recognized. The energy index gives the numerical value of an energy label, which could be observed in Table 7 in Appendix A with the proportions of the different energy labels in the dataset. Furthermore, a registration year for the current energy labels is noted, so the energy labels are linked to valid years of the labels. There is no registration year noted for the previous energy labels of the retail properties. So, it is assumed that the previous applicable label concerns all years before the year of registration of the new energy label. Two dummies are generated for the energy label variable in order to make a distinction between energy efficient and energy inefficient buildings. The five different classes under energy label A, label B, and C are classified as low energy consumption, and energy labels D, E, F, and G as high energy consumption (Kok and Jennen, 2012).

Demography

In the population numbers, CBS only includes people who are registered in the population register of a Dutch municipality. In principle, everyone who lives in the Netherlands for an indefinite period is included in the population register of the municipality of residence. People for whom no permanent residence can be designated are included in the population register of the municipality of The Hague. Not included are illegal people residing in the Netherlands and people subject to exceptional rules, for example diplomats and NATO military (Leeuwen, 2019). The distribution of the demography indicators gender and age groups follow from the population numbers. The population density is the population divided by the area of land in km^2. Private households consist of one or more residents who live alone or together in a living space and who themselves provide for their daily maintenance. In addition to single-person households, there are multi-person households.

Economy

The control variable economy in the research represents consumer confidence. Consumer confidence is an indicator that provides information about consumers' confidence and views on developments in the Dutch economy and their financial situation. Together with the sub-indicators economic climate and willingness to buy, these contribute to the prediction of short-term fluctuations in private consumption. The 'Consumenten Conjunctuuronderzoek' (CCO) measures the consumer confidence (CBS, 2021). The survey description is as follows: every month several questions are asked to approximately 1000 respondents. The questions are about the economic situation of the Netherlands and the financial situation of the respondents, both about the situation in the past and expectations for the future. Estimates of the percentages of positive and negative answers are published every month. Consumer confidence is derived from the answers to five specific questions which could be observed in Appendix A. The indicators can take a value from -100 (everyone answers negatively) to +100 (everyone answers positively). With a value of 0, the share of pessimists is equal to the share of optimists. The indicators have been adjusted for seasonality and sample noise.

E-commerce

This control variable contains figures on the turnover developments of internet sales of the retail sector. CBS uses the so-called 'Standaard Bedrijfsindeling' (SBI, 2008) for the classification of companies by main activity (KvK, 2020). The retail sector in this case has division 47. The data could be divided into different groups of shops: shops that mainly trade online and that mainly trade via other sales channels such as a physical shop or on the market. The development is shown as a percentage change compared to a previous period and by means of indices with 2015 as the base year. The measured turnover development from internet sales relates to retail companies in medium and large companies, which represents companies with 10 employees or more. These companies represent approximately 65-70% of online retail sales. Small businesses were not included. This could cause problems for the representativeness of this research. In addition, not all years are available since 2015 is measured as the base year and before that no data is available on these e-commerce factors. Moreover, e-commerce is measured on a national level which makes it almost impossible to control for these effects. These reasons have caused the exclusion of the e-commerce variables from the empirical model. However, it seems a crucial factor to take into account when analyzing retail vacancy, although only when all years are available and is on microlevel.

3.4 Descriptive statistics

The descriptive statistics of all variables that are ultimately included in both the simple logit regressions and the panel logit regressions could be found in Table 8. In the upper part of the table, the dataset for the simple logit regressions takes into account the data as cross-sectional data, whereas in the lower part the data for the panel logit regressions takes into account the changes over time. This table provides an overview of all variables and the most important values including the number of observations, mean, standard deviation, and the minimum and maximum value. Two times two models are integrated into Table 8, in Model 1 and 3 the population density variable on urbanity level and the control variable of demography are based on postcode4 level, and in Model 2 and 4 these variables are based on municipality level. All the other variables are based on the same level for both models.

Table 8. *Summary statistics all variables*

Descriptive Statistics Cross-sectional data			**Model 1 Postcode4 level**					**Model 2 Municipal -ity level**		
Variable + scale	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Dependent variable										
retail vacancy	707211	.053	.223	0	1	706794	.053	.223	0	1

(unit level)										
Independent variables										
Store agglomeration										
(p4 and p6 level)										
ln number stores p4	707211	5.110	1.110	0	8.888	706794	5.110	1.110	0	8.888
ln number stores p6	707211	1.678	1.023	0	6.326	706794	1.677	1.023	0	6.326
Population density										
(Model 1: p4 level,										
Model 2: municipality)										
urbanity level 1	449800	.567	.496	0	1	679814	.408	.493	0	1
urbanity level 2	449800	.283	.451	0	1	679814	.378	.484	0	
urbanity level 3	449800	.100	.301	0	1	679814	.149	.354	0	1
urbanity level 4	449800	.040	.197	0	1	679814	.060	.232	0	1
urbanity level 5	449800	.009	.096	0	1	679814	.005	.069	0	1
Energy efficiency										1
(unit level)										
lowenergy	707211	.236	.425	0	1	706794	.236	.425	0	
highenergy	707211	.154	.361	0	1	706794	.154	.361	0	1
										1
Control variables										
ln floorspace	707209	5.062	1.266	0	13.816	706791	5.061	1.266	0	13.816
(unit level)										
Demography										
(p4 level)										
inh upto 15years	697270	1226.441	773.919	0	6880	696901	1227.028	774.357	0	6880
inh 15to25years	697270	1269.568	824.905	0	4940	696901	1270.254	825.445	0	4940
inh 25to45years	697270	2835.568	1655.97	0	9870	696901	2837.083	1657.183	0	9870
inh 45to65years	697270	2254.921	1099.132	0	8920	696901	2255.505	1099.302	0	8920
inh 65yearsandolder	697270	1465.731	785.651	0	5370	696901	1465.745	785.457	0	5370
populationdensity ~h	584133	2730.98	1922.289	121	6620	678453	2600.419	1934.795	83	6620
privatehouseholds n	537005	143000.8	148296.5	4808	470223	678453	134103.2	145521.3	4805	475368
householdsize avg	537005	2.056	.192	1.64	2.71	678453	2.064	.200	1.6	2.7
Economy										
(national level)										
economicsituat~xt12m	697316	1.403	37.518	-59	41	696947	1.394	37.519	-59	41
financialsitua~st12m	697316	-12.600	8.774	-27	1	696947	-12.602	8.773	-27	1

Descriptive Statistics Panel data			**Model 3 Postcode4 level**					**Model 4 Municipal -ity level**		
Variable + scale + years	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Dependent variable										
retail vacancy	779723	.098	.297	0	1	779723	.098	.297	0	1
(unit level) (2012-2020)										
Independent variables										
Store agglomeration										
(p4 level: 1200 unique values) (2012-2020)										
(p6 level: 34606 unique values) (2012-2020)										
ln number stores p4	779723	5.117	1.109	0	8.888	779723	5.117	1.109	0	8.888
ln number stores p6	779723	1.687	1.015	0	6.326	779723	1.687	1.015	0	6.326
Population density										
(Model 3: p4 level) (2014-2019)										
(Model 4: municipality: 139 unique observations) (2012-2020)										
urbanity level 1	505055	.562	.496	0	1	749615	.408	.492	0	1
urbanity level 2	505055	.288	.453	0	1	749615	.378	.485	0	

urbanity level 3	505055	.101	.302	0	1	749615	.149	.356	0	
urbanity level 4	505055	.039	.195	0	1	749615	.060	.237	0	1
urbanity level 5	505055	.009	.094	0	1	749615	.005	.069	0	1
<u>Energy efficiency</u>										1
(unit level) (2012-2020)										1
lowenergy	779723	.234	.423	0	1	779723	.234	.423	0	
highenergy	779723	.150	.357	0	1	779723	.149	.357	0	
										1
										1
Control variables										
ln floorspace	779715	5.045	1.241	0	13.816	779714	5.045	1.240	0	13.816
(unit level) (2012-2020)										
<u>Demography</u>										
(unit level) (2012-2020)										
inh upto 15years	766226	1218.857	772.059	0	6880	766278	1218.829	772.129	0	6880
inh 15to25years	766226	1269.496	829.064	0	4940	766278	1269.536	829.118	0	4940
inh 25to45years	766226	2821.137	1649.526	0	9870	766278	2821.232	1649.794	0	9870
inh 45to65years	766226	2249.327	1093.373	0	8920	766278	2249.324	1093.506	0	8920
inh 65yearsandolder	766226	1477.227	786.859	0	5370	766278	1477.189	786.874	0	5370
populationdensity ~h	643519	2667.995	1897.325	121	6620	748254	2547.904	1916.92	83	6620
privatehouseholds n	588329	139688.94	147040.01	4808	470223	748254	130820.47	143886.14	4805	475368
householdsize avg	588329	2.057	.192	1.64	2.71	748254	2.066	.199	1.6	2.7
<u>Economy</u>										
(national level: 1 unique observation per year) (2012-2020)										
economicsituat~xt12m	766272	2.58	36.869	-59	41	766324	2.579	36.868	-59	41
financialsitua~st12m	766272	-12.261	8.808	-27	1	766324	-12.261	8.808	-27	1

3.5 <u>Methodology</u>

To test the research hypotheses, a statistical analysis is performed using multiple regressions. Based on these regressions, the extent to which the independent variables are predictors for the dependent variable retail vacancy will be investigated. Therefore, the purpose of the regression models is to measure whether the independent variables are significant predictors for retail vacancy. This takes into account and is corrected for differences in factors of demography, economy, and retail floorspace measures over time. To make a thorough analysis over time from 2012-2020 with retail vacancy as the dependent variable, a discrete choice model could be chosen. Here, the dependent variable y is a binary variable that has either the value 0 which means that there is no vacancy, or the value 1 when there is vacancy. One option is to choose the linear probability model to estimate the probability of vacancy per variable. The ordinary least squares (OLS) econometric model will look like the following equation:

$$yi = \beta 0 + \beta 1xi,1 + \beta 2xi,2 + \cdots + \beta kxi,k + \varepsilon \qquad (1)$$

And the linear probability model has the following linear function:

$$E(yi|xi) = \beta 0 + \beta 1xi,1 + \beta 2xi,2 + \cdots + \beta kxi,k + \varepsilon \qquad (2)$$

where β is the increment in probability which will be associated with a one-unit increase in that particular explanatory factor x. If the explanatory variables are exogenous, the beta parameters will be estimated by doing an OLS estimation that is consistent and unbiased. Binary outcome models estimate the probability that y = 1 as a function of the independent variables. The conditional expectation is:

$$E(yi|xi) = \mathrm{P}\ (yi = 0|xi) * 0 + \mathrm{P}\ (yi = 1|xi) * 1 \tag{3}$$
$$E(yi|xi) = P\ (yi = 1|xi)$$

Problems that could occur with the linear probability model are the following (DeMaris, 1995); Firstly, linear functions are always unbounded both upwards and downwards. This means that it allows the predicted probabilities to be outside the range which they normally accept to be in with the binary dependent variable in this model; so, the [0,1] range. Their interpretation as probabilities is somewhat non-sensical, because there are no restrictions for the probabilities to be bounded by the [0,1] interval. That is a problem in the assumption; when the probabilities are estimated, the results will not be restricted to [0,1]. A solution could be whenever the estimated probabilities are greater than 1 just set them to 1, and when probabilities are below 0 it is common to set them to 0. Making these changes will not hurt in large samples. Secondly, heteroscedasticity, the linear probability model is not the most efficient unbiased estimator of the parameter's beta. The error term ε will be heteroscedastic. This means that the variance of the error term varies from one observation to another. To take care of that robust standard errors could be used, or alternatively a generalized least squares procedure could be applied. Thirdly, non-normality, because of the discrete nature of the errors some small issues of using normal distributions for inference small samples could be run into. There is a non-normal distribution here. The outcome y only takes two values, implying that the error term also takes only two values, so that the usual 'bell-shaped' curve describing the distribution of errors does not hold. Lastly, linear models assume a constant marginal effect. However, with a dummy variable as dependent variable, this becomes unrealistic. To overcome these problems, the non-linear logit model or the probit model could be considered (Hill et al., 2018). The interpretation of a binary logit or probit model looks like the following equation with a latent variable:

$$yi' = \beta 0 + \beta 1xi,1 + \beta 2xi,2 + \cdots + \beta kxi,k + \varepsilon \tag{4}$$
$$yi = I\{yi' > 0\} = \begin{cases} 1\ if\ yi' > 0 \\ 0\ if\ yi' \leq 0 \end{cases} \tag{5}$$

where ε is an error term with a standard logistic or normal distribution which is independent of xi. This error term is symmetric around zero and continuous. $I\{...\}$ is an indicator function which is one if the event in {…} is true, and zero otherwise. The purpose of a logit and probit model is to estimate the odds that an observation with particular characteristics will fall into one of the specific categories. So, in this case it estimates the odds of the chance of retail vacancy. The logit and the probit model produce almost identical results, different coefficients but similar marginal effects (Hsiao, 1996). Because of this indifference, the logit model is chosen for this research. The logistic regression equation has the following form:

$$ln\left(\frac{P}{1-P}\right) = \beta 0 + \beta 1xi,1 + \beta 2xi,2 + \ldots + \beta kxi,k + \varepsilon \qquad (6)$$

$$Odds = \frac{P}{1-P} = e^{(\beta 0 + \beta 1xi,1 + \cdots + \varepsilon)} \qquad (7)$$

ln (P/(1-P)) is the natural logarithm of the odds that y = 1. The type of statements that could be made based on this applied model include statements about the estimation of the odds favoring that y = 1. So, if $x1$ increases with one unit, ln(odds) will increase with $\beta 1$. Stating it differently, if $x1$ increases with one unit, odds will multiply with exp ($\beta 1$) (DeMaris, 1995). This means, a one unit increase in $x1$ increases the odds of y = 1 with (exp ($\beta 1$) – 1) *100% times, compared to y = 0 and keeping all other variables constant.

The logistic regression assumptions are somewhat different from the ones of linear regression. Firstly, the dependent variable is binary or dichotomous; it only takes on two possible outcomes. Secondly, the observations are independent of each other. Thirdly, the model is correctly specified if there is linearity between the logit of the dependent variable y and each independent variable x. Fourthly, absence of multicollinearity must be met. So, there are no high correlations among the explanatory variables. Fifthly, absence of influential observations must be taken care of. This means there are no extreme values or outliers in the dataset (Stoltzfus, 2011). Lastly, the sample size is sufficiently large to draw valid conclusions. In comparison with linear regression, logistic regression does not require linearity between the dependent variable and the explanatory variables. Furthermore, the residuals are not required to be normally distributed or to have constant variance which is called homoscedasticity.

Besides, when estimating logit models, it is of great importance to investigate the goodness of fit of the model. This is done by calculating the percent correctly predicted values. The percent correctly predicted values are the proportion of true predictions to the total predictions. The Pseudo R^2 and the maximum likelihood estimates could be calculated as well to investigate the goodness of fit of the model (DeMaris, 1995). It compares the unrestricted log-likelihood for the model that will be estimated and the restricted log-likelihood with only an intercept. If the independent variables have no explanatory power, the restricted model will be the same as the unrestricted model and the R-squared will be zero. The formula of the Likelihood Ratio (LR) test is the following:

$$LR = -2\ (ln\ L_{restricted} - ln\ L_{unrestricted}) \qquad (8)$$

3.6 Model building

The strategy used in this research to build a suitable model is the step-up method by starting with very few variables and then adding increasingly more variables to the model (Stoltzfus, 2011). A suitable model should contain all explanatory variables x that are of essence in explaining the dependent variable y. The logit model has been compiled based on the variables that emerged from academic literature in chapter 2 and the variables available in the dataset. Firstly, the independent variables for store agglomeration and the control variable floorspace are log-transformed since these variables in the original continuous data format do not follow a bell-curved distribution but are right-skewed. Transforming these data removes the skewness of the original data and thus creates a more normal distribution to give the statistical analysis results a more valid meaning, the transformation of the variable 'number of stores on postcode4 level' is presented in Figure 3 in Appendix B.

To test for the assumptions of logistic regression, first of all it is crucial that the dependent variable is binary. In this research the dependent variable takes only two possible values, so the assumption holds. Secondly, the observations need to be independent of each other. Testing for independence is done by plotting the residuals against the time in years. In Figure 4 in Appendix C it could be observed that the residuals are randomly distributed over time, this suggests indeed independence. The third assumption indicates that there are no extreme outliers or influential values in the dataset. In Figure 5 it could be observed that there are indeed no extreme outliers visible by plotting the standardized Pearson residuals against the predicted probabilities. Fourthly, linearity needs to be tested which could be done with the 'lowess' command in Stata. However, due to the extremely large dataset Stata was unable to run the test. Another way is by executing the Box-Tidwell test. The Box-Tidwell regression model has been run on the three main independent variables that are tested in the hypotheses. The outcome of the Box-Tidwell test concludes that the transformed variables of the number of stores in a postcode 4 and postcode 6 area do not have a linear relationship with the logit of the dependent variable retail vacancy. This could be observed in Table 9 in Appendix C, at the bottom of the regression table. The test of nonlinearity of the variable 'ln numberstores p4' and 'ln numberstores p6' are statistically significant with a p-value = 0.004 and 0.000 respectively. When the explanatory variables were linear to the logit of the dependent variable, p1 should be equal to 1. Actually, the p1 values are equal to 1.74 and -1.02. So, this assumption does not hold. It may become more linear when creating an interaction term between the log-transformed variables and the variables before the transformation of the number of stores in postcode4 and 6. Fifthly, the assumption must hold that the sample size is large enough to draw valid conclusions from the logistic regression models. The dataset is indeed sufficiently large. Lastly, the model is created by looking at the correlations between the various variables and by conducting Variable Inflation Factors (VIF) tests to detect multicollinearity between the explanatory variables, which is as well a logistic regression assumption. A VIF score states the strength of the correlation between the explanatory variables. The higher the value of VIF, the higher the multicollinearity with the particular independent variable. A VIF exceeding a value of 10 is problematic and indicates high multicollinearity (Midi et al, 2010). Multicollinearity could be problematic in a regression model since determining individual effects of the independent variables might become less reliable, this will cause problems in terms of interpretability. Multicollinearity could be fixed by iteratively dropping variables starting with the variable indicating the largest VIF value. Consequently, VIF values of other variables will reduce to a varying extent. A considerable amount of demography variables is not included in the model since these approximately all explained the same phenomenon. The final VIF scores with all desired values are presented in Table 10a in Appendix C. The value of R^2 for the overall regression for the various subset models slightly increased after removing the correlated predictors. The McFadden Pseudo R^2 is used to find out how precisely the binary logistic regression model explains and fits the observed data. The coefficient of determination Pseudo R^2 that is bounded between [0,1] yields a global check on the regression model. The higher Pseudo R^2 the more variation the model explains. The highest obtainable Pseudo R^2 on the final models on municipality level is extremely low for the simple logit model; 0.068. This means that only 6.8% of the variance of the dependent variable could be explained by the explanatory variables. On postcode4 level the highest obtainable Pseudo R^2 is 0.087. These models excluded the variables 'population growth relative', 'high income' and the e-commerce variables, since when these variables were added to the model, the Pseudo R^2 decreased. So, the Pseudo R^2 is low because of the possibility and suspicion of omitted variable bias. When comparing de Pseudo R^2 with other scientific studies, the R^2 in this research is slightly lower (Teulings et al., 2018; DeMaris, 1995). And the VIF scores all remained with desirable values with a mean VIF score of 2.81, which could be observed in Table 10b in Appendix C.

Detecting a specification error, the 'linktest' in Stata could be issued immediately after a regression. It runs the linear predicted value (_hat) and the linear predicted value squared (_hatsq) as predictors to rebuild the model, which could be observed in Table 11 in Appendix D. The predicted value from the model is the linear predicted value and is statistically significant since the model is not completely incorrectly specified. The linear predicted value squared is significant as well which it should not be. This implies that relevant variables are omitted from the model. This is very likely true, according to academic literature e-commerce has a substantial effect on retail vacancy. When adding the e-commerce variables to the model and checking the significance of the linear predicted value squared, this value becomes insignificant. So, it could indeed be concluded that e-commerce is a variable that is omitted from the model that indeed plays an important role in the drivers of retail vacancy. E-commerce is not included in the created model, since the quality of the model decreases otherwise. The cause could be that considerably few observations of e-commerce are present in the dataset. Besides e-commerce variables, other property-specific characteristics may have an influence on retail vacancy of which no data is available. So, these crucial variables were all not included in the model and consequently lead to omitted variable bias.

Based on the created model, it is tested which variables have an influence on the occurrence of retail vacancy. Where the natural logarithm of the odds that a retail property becomes vacant is a function of the independent variables and the control variables mentioned in the previous chapter.

4. RESULTS

Here, the outcomes of the quantitative analysis of the logit model will be presented to test the three aforementioned hypotheses of this research. Firstly, the analysis will be performed with simple logit models, in which the panel structure is being ignored and thus treated as cross-sectional data. For the simple logit analysis retail properties are deleted from the dataset after the first-time vacancy is observed for the same property over time. This means no fluctuations between occupancy and vacancy could be observed for single properties over time. Thereafter, the analysis will present the results of an 'xtlogit' model, which will take into account the panel data structure. This analysis is done on the data where no buildings are deleted after vacancy is observed. Furthermore, these results will be discussed using academic literature.

4.1 Findings simple logit model

The results of the binary logistic regressions on municipality level are presented in Table 12 and provide insights into the relationship of the dependent retail vacancy variable, the independent variables, and the significant relationships of the control variables. Moreover, year fixed effects are included to control for observable and unobservable factors changing each year. In Table 13 in Appendix D the results of the logistic regression based on the dataset linked on postcode4 level are presented. This dataset contains a substantial reduction in the number of observations, even after removing some variables from the model with too many missing values. Still, several control variables had a lower availability on postcode4 level to be linked to the dataset. Besides, significance changed for several variables included in the model.

Table 12. *Logistic regression (municipality level)*

	Model 1		Model 2		Model 3		Model 4		Model 5	
retail_vacancy	Coef.	St. Err.	Coef.	St.Err.	Coef.	St.Err.	Coef.	St.Err.	Coef.	St.Err.
ln_numberstores_p4	.031***	.008								
ln_numberstores_p6			.051***	.007	.052***	.007			.136***	.01
energy_efficient	-1.264***	.02	-1.28***	.02	-1.275***	.02	-1.264***	.02	-1.28***	.02
energy_inefficient	-1.696***	.029	-1.683***	.029	-1.684***	.029	-1.695***	.029	-1.676***	.029
inh_upto_15years	3.37x10^{-5}	2.09x10^{-5}	8.51x10^{-7}	1.79x10^{-5}	1.82x10^{-5}	1.76x10^{-5}	-4.99x10^{-6}	1.8x10^{-5}	-8.66x10^{-6}	1.8x10^{-5}
inh_15to25years	-1.87x10^{-5}	1.33x10^{-5}	-1.74x10^{-5}	1.33x10^{-5}	-1.34x10^{-5}	1.33x10^{-5}	-1.52x10^{-5}	1.33x10^{-5}	-1.86x10^{-5}	1.34x10^{-5}
inh_25to45years	5.33x10^{-5}***	1.06x10^{-5}	6.68x10^{-5}***	9.76x10^{-6}	7.56x10^{-5}***	9.65x10^{-6}	6.81x10^{-5}***	9.77x10^{-6}	7.35x10^{-5}***	9.79x10^{-6}
inh_45to65years	-1.659x10^{-4}***	1.79x10^{-5}	-1.589x10^{-4}***	1.78x10^{-5}	-1.756x10^{-4}***	1.75x10^{-5}	-1.644x10^{-4}***	1.79x10^{-5}	-1.56x10^{-4}***	1.78x10^{-5}
inh_65yearsandolder	1.947x10^{-4}***	1.38x10^{-5}	1.938x10^{-4}***	1.36x10^{-5}	1.905x10^{-4}***	1.34x10^{-5}	2.06x10^{-4}***	1.35x10^{-5}	1.925x10^{-4}***	1.36x10^{-5}
populationdensity_~h	-4.92x10^{-5}***	7.01x10^{-6}	-4.69x10^{-5}***	7.02x10^{-6}	-1.33x10^{-5}**	5.71x10^{-6}	-4.74x10^{-5}***	7.01x10^{-6}	-4.37x10^{-5}***	7.07x10^{-6}
privatehouseholds_n	-2.19x10^{-6}***	9.03x10^{-8}	-2.18x10^{-6}***	9.03x10^{-8}	-2.06x10^{-6}***	8.83x10^{-8}	-2.17x10^{-6}***	9.04x10^{-8}	-2.20x10^{-6}***	9.02x10^{-8}
householdsize_avg	.016	.057	.037	.056	-.054	.052	.06	.056	.05	.056
economicsituat~xt12m	-.015***	.002	-.015***	.002	-.014***	.002	-.015***	.002	-.014***	.002
financialsitua~st12m	.024***	.01	.025**	.01	.025**	.01	.027***	.01	.023**	.01
lnfloorspace	-.29***	.005	-.288***	.006	-.286***	.005	-.291***	.005	-.286***	.005
urbanitylevel										
2.urbanitylevel	-.192***	.024	-.188***	.024			-.187***	.024	.064*	.037
3.urbanitylevel	-.266***	.033	-.256***	.033			-.26***	.033	-.002	.049
4.urbanitylevel	-.248***	.041	-.23***	.041			-.243***	.041	-.011	.064
5.urbanitylevel	-.996***	.108	-.97***	.108			-.979***	.108	-.712***	.197
ln_numberstores_p6 x 2.urbanitylevel									-.141***	.015
ln_numberstores_p6 x 3.urbanitylevel									-.139***	.019
ln_numberstores_p6 x 4.urbanitylevel									-.121***	.029
ln_numberstores_p6 x 5.urbanitylevel									-.144	.101
year										
2013	-.33***	.069	-.327***	.069	-.333***	.069	-.309***	.069	-.338***	.069
2014	1.088***	.322	1.106***	.32	1.077***	.32	1.197***	.321	1.049***	.321
2015	.659***	.166	.669***	.166	.657***	.166	.718***	.166	.642***	.166
2016	.511***	.106	.516***	.105	.509***	.105	.55***	.105	.501***	.105
2017	.461***	.072	.463***	.072	.449***	.072	.49***	.072	.456***	.072
2018	.575***	.058	.575***	.058	.562***	.058	.596***	.058	.572***	.058
2019 (omitted)										
2020 (omitted)										
constant	-.976***	.131	-.976***	.131	-1.033***	.127	-.929***	.13	-1.193***	.133

number of obs.	668,583	668,583	668,583	668,583	668,583
chi2	12448.360	12573.988	12500.448	12412.813	12658.057
prob > chi2	0.000	0.000	0.000	0.000	0.000
pseudo R2	0.067	0.067	0.067	0.067	0.068

*Note: The dependent variable is the binary variable retail_vacancy, indicating whether a retail property is vacant or not vacant. *, **, *** are significant at 10%, 5% and 1% respectively. The standard errors are clustered on ID.*

The first hypothesis looks at the association of the number of stores in the same postcode4- and postcode6 area. With the data on municipality level, the association of both the number of stores in the same postcode4- and postcode6 area, in model 1 and model 2 respectively, turn out to be significantly different from zero at the 99% confidence level. In addition, it both indicates a small positive association with retail vacancy. The interpretation with the log-transformed variable becomes that each k-fold increase in x is associated with a change in the odds by a multiplicative factor of k^{β}. So, a 1% increase in the number of stores in a postcode4 area increases the odds of a store becoming vacant by a factor of $k^{\beta} = 1.01^{0.031} = 1.00031$ times. Or explained otherwise, if the number of stores in the same postcode4 area increases with 1%, the odds ratio of a store becoming vacant increases with 0.03%, keeping all other explanatory variables constant. And a 1% increase in the number of stores in a postcode6 area leads to a change in the odds of a store becoming vacant by a factor of $k^{\beta} = 1.01^{0.051} = 1.00051$ times. So, if the number of stores in the same postcode6 area increases with 1%, the odds ratio of a store becoming vacant increases with 0.05%, keeping all other explanatory variables constant. It could be concluded that both the increase of the number of stores in a postcode4- and a postcode6 area hardly have an association with retail vacancy. Since the economic effect size seems incredibly small, a 10% increase in the number of stores in a postcode4- and postcode6 area will be investigated. A 10% increase in the number of stores in a postcode4 area increases the odds of a store becoming vacant by a factor of $k^{\beta} = 1.1^{0.031} = 1.0030$ times. Or explained otherwise, if the number of stores in the same postcode4 area increases with 10%, the odds ratio of a store becoming vacant increases with 0.3%, keeping all other explanatory variables constant. And a 10% increase in the number of stores in a postcode6 area leads to a change in the odds of a store becoming vacant by a factor of $k^{\beta} = 1.1^{0.051} = 1.0049$ times. Explained otherwise, if the number of stores in the same postcode6 area increases with 10%, the odds ratio of a store becoming vacant increases with 0.5%, keeping all other explanatory variables constant. Analyzing these results, the associations are in opposite direction as expected. Namely, as the number of stores increases, the odds of a store becoming vacant also slightly increases. So, the theoretical hypothesis formulated in Chapter 2 cannot be supported.

The results of Model 3 regarding the store agglomeration variable 'ln_numberstores_p6' by leaving out the urbanity level variable is shown in Table 12 and in Model 4 this is done vice versa. It could be observed that there is not a clear difference in association of these particular variables when leaving out the one or the other. In Model 5 the interaction term between the aforementioned variables is included. It could be concluded that the interaction shows for most levels significant interaction results, only the interaction with urbanity level 5 turns out to be insignificant. In the different urbanity levels 1 till 4 there exists a different relationship between the number of stores in a postcode6 area with the odds of becoming vacant. The slopes of the regression lines between the number of stores in a postcode6 area and retail vacancy are different for the different categories of urbanity level.

For the analysis on the data on postcode4 level, the variables with a considerably lower number of observations are left out of the analysis. The applicable variables are urbanity level, private households, and household size. Only one of the two coefficients on store agglomeration in the dataset on postcode4 level turn out to be significantly different from zero. For the number of stores in a postcode4 area, the coefficient seems not significantly different from zero and is incredibly small. And the coefficient for the number of stores in a postcode6 area is significantly different from zero at the 99% confidence interval. This could be observed in Table 13 in Model 1 and 2 respectively. A 10% increase in the number of stores in a postcode6 area leads to a change in the odds of a store becoming vacant by a factor of $k^{\beta} = 1.1^{0.085} = 1.00813$ times, holding all other explanatory variables constant. To conclude, the tiny positive association of the number of stores in a postcode4 area turns out to be insignificant. Besides, for the number of stores in a postcode6 area, the theoretical hypothesis could not be supported. As the association of the number of stores in a postcode6 area is slightly bigger than in the previous analysis with data on municipality level but still very small and has a positive association in relation to retail vacancy which is contrary to the expectation.

The difference in association between the number of stores in a postcode4 area and a postcode6 area could be explained by the difference in size of the concerning area types. A postcode6 area refers to a territory consisting of approximately 25 properties. Whereas a postcode4 area refers to a small hometown or neighborhood within a large city and has a surface ranging between 1.1 km^2 and 8.3 km^2 in the Netherlands. So, no conclusion could be drawn on postcode6 level, since in most of these areas there are not even retail locations, leave alone a lot. Consequently, it becomes hard to test the association. In most postcode4 areas more diverse retail establishments could be found. Sevtsuk (2014) states that when the customer purchases everyday products frequency increases, the distance between stores decreases and the density between retailers is higher. Higher store densities are found between retailers that sell frequently purchased goods such as food and drinks compared to infrequently purchased goods such as furniture. It is thus really dependent on the store category. According to their specific products and clients, retailers carefully pick locations. From this research, no conclusion could be drawn based on the different store categories. Besides, as the density of customers rises, the density of stores also increases. The distribution of retailers is affected by the accessibility to form attractive locations.

The second hypothesis investigates the association of the population density measured in residential addresses per km^2 categorized in urbanity levels from high number of addresses per km^2 (urbanity level 1) to low number of addresses per km^2 (urbanity level 5). The prediction entails the higher the urbanity level the less likely it is that a store becomes vacant. The highest urbanity level 1 is taken as the reference category. As could be observed in Table 12 in Model 1, 2, and 4 is that all urbanity level coefficients are significantly different from zero at the 99% confidence level and these three different models present almost identical coefficients. In Model 4 the coefficient of urbanity level 2 shows that the odds of a store becoming vacant decreases with ((exp $^{(-0.187)}$ -1) *100) % = 17.06% than a store becoming vacant in urbanity level 1, compared to a store not becoming vacant and keeping all other variables constant. In urbanity level 5 the odds of a store becoming vacant decreases with ((exp $^{(-0.979)}$ -1) *100) % = 62.43% than a store becoming vacant in urbanity level 1, compared to a store not becoming vacant and keeping all other variables constant. The odds of a store becoming vacant decreases more and more as the urbanity level decreases. However, only the association of urbanity level 4 shows a slightly lower negative association compared to urbanity level 3, as could be observed in Table 12 in Model 1, 2, and 4. Overall, it could be stated that in lower urbanity level areas it is less likely a store becomes vacant. The odds of a store becoming vacant is less likely in areas where there are fewer addresses per km^2. This is contrary to the hypothesis. It may as well be due to the fact that in urbanity level 5 there are very few retail locations, compared to urbanity level 1. So, the theoretical hypothesis could not be supported. Hollander et al. (2018) show contrary insights to the findings of this research, where population decline came hand in hand with increases in vacancy. The contrary outcomes in comparison with scientific literature on this topic are most likely due to the omitted variable bias that is present in the model due to unavailability of some essential variables that have a contribution in explaining retail vacancy as well.

The third and last hypothesis focuses on the association of building energy efficiency on retail vacancy. On municipality level in all 5 models in Table 12, both the categories on building energy efficiency have a negative sign and are significantly different from zero at the 99% confidence level. The reference category is properties not having an energy label at all. In Model 1 if the property has a high energy label, so a building that is energy efficient with low energy emissions, the odds of that retail building becoming vacant will multiply with exp $^{(-1.264)}$ = 0.2825 or the odds of that retail building becoming vacant decreases with ((exp $^{(-1.264)}$ -1) *100) % = 71.75% than a property without having an energy label at all, keeping all other explanatory variables constant. If the property has a very low energy label, so a building that is not energy efficient and has high energy emissions, the odds of that retail building becoming vacant will multiply with exp $^{(-1.696)}$ = 0.1834 or the odds of that retail building becoming vacant decreases with ((exp $^{(-1.696)}$ -1) *100) % = 81.66% than a property without having an energy label at all, holding all other variables constant. It could be concluded that buildings having an energy label are less likely to become vacant, compared to retail buildings without an energy label.

When splitting the energy labels into three groups instead of two, distinctions in more detail could be realized between the different energy label ranks. Where the most energy efficient labels enclose all 5 ranks under energy label A, moderate energy efficiency encloses the labels B, C, and D, and energy inefficient labels include energy labels E, F, and G. From these three groups it could be concluded that having an energy label has indeed a negative association with retail vacancy. However, there does not appear a clear distinction in association between the different energy label heights as could be observed in Table 14 in Appendix D. Thus, the rank of an energy label does not differ in association necessarily. There is no clear growing probability visible when observing the height of the energy label and the chance of retail vacancy. This finding is in line with the results of Fuerst and McAllister (2011) who stated that a superior score in energy efficient buildings could have an effect on the financial performance of the property. Cost reductions are associated with these higher financial performances, such as lower operating expenses and lower vacancy rates. However, the authors highlight as well that there is hardly evidence on the extent of these gains. This is somewhat equivalent to the conclusion that could be drawn from this research. The extent of the benefits in having less chance that a retail property becomes vacant could not be matched to the height of the energy label, only on having a label. So, the theoretical hypothesis could not be supported. Apparently, there is no evidence from the analysis that the more energy efficient a retail property is, the lower the odds of retail vacancy. Perhaps this result is not in line with the expectation since other important factors of retail properties individually play an even important role in explaining retail vacancy as the energy label of a building. However, these factors could not be included in the model, so retail vacancy could not be explained at best.

The regression outcomes on postcode4 level related to the building energy efficiency show similar results compared to the analysis on municipality level. The coefficients show a slightly higher association with retail vacancy. However, the same pattern could be observed as the height of the label does not make a difference for the effect on retail vacancy only owing to the fact that they have a label. Correspondingly, it shows here that buildings with an energy label are less likely of becoming vacant, compared to retail buildings without an energy label.

A robustness analysis is done to analyze how the effects of building- and location-characteristics on retail vacancy differ between different residences with the three different shopping area types with inner-city or main shopping area. These different types are described in Chapter 3. It concerns the following three shopping area categories: inner cities containing more than 400 stores (area type 1), large main shopping areas containing 200-400 stores (area type 2), and small main shopping areas containing 100-200 stores (area type 3). Six residences per category are picked out of the data to analyze the heterogeneity between the results. As could be observed in Table 15 in Appendix D, area type 1 and 2 have a negative association with retail vacancy and area type 3 a positive association with retail vacancy. As could be observed as well is that the different area types show insignificant results, only area type 2 show significant results when adopting area type 1 as the reference category. All other coefficients in the model show similar coefficient outcomes compared to the coefficients in Table 12 Model 2.

4.2 Findings panel logit model

For this model, the same variables are log-transformed as in the simple logit models. The VIF scores are checked and are acceptable, with a mean VIF score of 2.81, for all same variables included in the simple logit model on municipality level. So, the same variables are used for running the logistic regression with the panel structure taken into account. The Hausman test is considered to choose between fixed-effects or random-effects to include in the model. However, the test results indicated that there may be problems computing the test. So, the decision is not based on this test. In the end, the random-effects estimator is chosen here as an option, since it is the default and because the features of random-effects are able to investigate time-invariant causes of the dependent variable retail vacancy and features of fixed-effects are not able to analyze these. An example of a time-invariant explanatory variable that must be considered in this research is ‘lnfloorspace’. The ‘lrmodel’ option is added to perform the likelihood-ratio model test. The estimated results of the random-effects logistic regression could be observed in Table 16 below. This type of regression is only executed on the data matched on municipality level. Additional to the simple logit regression output, this output includes the panel-level variance component, which is measured as the log of the variance. Besides, at the bottom of the output table, a likelihood-ratio test is performed. This test compares the pooled estimator (logit) with the panel estimator.

Table 16. *Random-effects logistic regression (municipality level)*

	Model 1		**Model 2**	
retail_vacancy	Coef.	St.Err.	Coef.	St.Err.
ln_numberstores_p4	-.475***	.023		
ln_numberstores_p6			.113***	.018
energy_efficient	-6.427***	.063	-6.426***	.062
energy_inefficient	-6.382***	.077	-6.351***	.077
inh_upto_15years	-.001***	$5.94x10^{-5}$	$-1.948x10^{-4}$***	$5.34x10^{-5}$
inh_15to25years	$-2.99x10^{-5}$	$4.03x10^{-5}$	$-7.36x10^{-5}$*	$4.01x10^{-5}$
inh_25to45years	$3.106x10^{-4}$***	$3.01x10^{-5}$	$8.64x10^{-5}$***	$2.83x10^{-5}$
inh_45to65years	$-4.166x10^{-4}$***	$5.17x10^{-5}$	$-4.171x10^{-4}$***	$5.18x10^{-5}$
inh_65yearsandolder	.001***	$4.23x10^{-5}$	.001***	$4.17x10^{-5}$
populationdensity_~h	$4.77x10^{-5}$**	$1.98x10^{-5}$	$3.14x10^{-5}$	$1.97x10^{-5}$
privatehouseholds_n	$-8.71x10^{-6}$***	$2.75x10^{-7}$	$-8.83x10^{-6}$***	$2.74x10^{-7}$
householdsize_avg	.458***	.17	-.149	.168
economicsituat~xt12m	.002***	$2.84x10^{-4}$	.003***	$2.829x10^{-4}$
financialsitua~st12m	.04***	.001	.046***	.001
urbanitylevel				
2.urbanitylevel	.155**	.061	.137**	.061
3.urbanitylevel	-.382***	.085	-.437***	.085
4.urbanitylevel	-.068	.114	-.068	.113
5.urbanitylevel	-2.005***	.265	-2.083***	.264
lnfloorspace	-2.153***	.018	-2.141***	.018
constant	4.619***	.376	3.63***	.374
lnsig2u	4.626	.009	4.615	.009
sigma u	10.103	.045	10.048	.045
rho	.969	.000	.968	.000

number of obs.	734851	734851
chi2	45731.263	45474.081
prob > chi2	0.000	0.000
LR test of rho=0: chibar2(01) = 1.8e+05 Prob >= chibar2 = 0.000		

*Note: The dependent variable is the binary variable retail_vacancy, indicating whether a retail property is vacant or not vacant. *, **, *** are significant at 10%, 5% and 1% respectively.*

The interpretation of the coefficients includes both the within-entity and between-entity effects. It represents the average association of X over Y when X changes across time and between buildings (IDs) by one unit. When comparing these results with the simple logit regression results, it could be observed that for most variables the estimates are larger in magnitude in the random-effects logistic regression. Besides, in Model 1 and Model 2 the significance for urbanity level 4 disappears. In addition, in Model 1 one of the age categories of inhabitants becomes insignificant, the rest of the variables are significantly different from zero. In Model 2 only the variables on population density and household size become insignificant. The highly significant likelihood-ratio test at the bottom of the regression output in Table 16, displays that it would not be appropriate to use the regular simple logistic regression instead. The likelihood-ratio test compares the pooled estimator (logit) with the panel estimator. Additionally, rho is included in the regression output, this describes the proportion of the total variance contributed by the panel-level variance component. For this research, approximately 97% of the variance could be explained by the panel structure of the analysis. So, it could be stated that the change of influences over time on retail vacancy is of great importance in explaining the drivers of retail vacancy. Furthermore, when taking the years into consideration in the regression analysis, most associations of the explanatory variables are of greater extent in relation to retail vacancy.

The first hypothesis tests for the association of the number of stores in a postcode4- and postcode6 area. A 10% increase in the number of stores in a postcode4 area leads to a change in the odds of a store becoming vacant by a factor of kβ = 1.1-0.475 = 0.9557 times, holding all other explanatory variables constant. And a 10% increase in the number of stores in a postcode6 area leads to a change in the odds of a store becoming vacant by a factor of kβ = 1.10.113 = 1.0108 times, keeping all other explanatory variables constant. The effect of the number of stores in a postcode4 area turns out differently than on the conclusion drawn in the previous analysis with the simple logit models. As the number of stores in a postcode4 area increases, the odds of a store becoming vacant slightly decreases. This is in line with the prediction made in chapter 2 grounded on academic literature. So, using the random-effects logistic regression, for the number of stores in a postcode4 area, the first theoretical hypothesis formulated in Chapter 2 can be supported.

The second hypothesis investigates the association of population density measured in urbanity level on retail vacancy. These associations all turn out to be significant, only the association of urbanity level 4 on retail vacancy compared to urbanity level 1 becomes insignificant. In Model 1 in urbanity level 2 the odds of a store becoming vacant increases with ((exp $^{(0.155)}$ -1) *100) % = 16.77% than a store becoming vacant in urbanity level 1, compared to a store not becoming vacant and keeping all other variables constant. This positive association with retail vacancy is in line with scientific literature and could not be concluded from the simple logit regression analysis. Except for Model 5 with the interaction term included, this model shows a positive association between urbanity level 2 and retail vacancy. This coefficient is significantly different from zero at the 90% confidence interval. There in urbanity level 2, the odds of a store becoming vacant increases with ((exp $^{(0.64)}$ -1) *100) % = 89.65% than a store becoming vacant in urbanity level 1, compared to a store not becoming vacant, when the variable 'ln_numberstores_p6' is equal to zero, and keeping all other variables constant. Urbanity levels 3 and 5 show similar results comparing them to the outcomes of the simple logit regressions.

Lastly, the third hypothesis investigates the association of building energy efficiency on retail vacancy. In Model 1, if the property has a high energy label, so a building that is energy efficient with low energy emissions, the odds of that retail building becoming vacant will multiply with exp $^{(-6.427)}$ = 0.0016 or the odds of that retail building becoming vacant decreases with ((exp $^{(-6.427)}$ -1) *100) % = 99.84% than a property without having an energy label at all, keeping all other explanatory variables constant. If the property has a very low energy label, so a building that is not energy efficient and has high energy emissions, the odds of that retail building becoming vacant will multiply with exp $^{(-6.382)}$ = 0.0017 or the odds of that retail building becoming vacant decreases with ((exp $^{(-6.382)}$ -1) *100) % = 99.83% than a property without having an energy label at all, holding all other variables constant. So, over time it becomes even clearer that having an energy label has a huge negative relationship on retail vacancy despite the rank the label has.

5. CONCLUSION & DISCUSSION

This section will summarize the main findings and provides an answer to the main research question and sub-questions. Thereafter, a critical assessment of the results and weaknesses regarding the data and methodology will be discussed. Then, the policy relevance will be stated, and recommendations will be proposed for future follow-up research. At the end, a critical reflection will be given on the research process.

5.1 Conclusion

This study aimed to investigate whether different particular building- and location-characteristics have an influence on retail vacancy in the Netherlands. The guiding main research question of this study is: *'What are the drivers of retail vacancy in the Netherlands?'* The theoretical framework has presented that e-commerce is a crucial factor having a substantial influence on retail vacancy. In addition, demographic changes, accessibility, floorspace, macro-economic changes, the imbalance between supply and demand, and Covid-19 all play a role in determining retail vacancy. Due to the almost unruly data, no clear conclusion could be drawn on most hypotheses. What could be concluded is that when using the random-effects logistic regression, the higher the number of stores in a postcode 4 area, the less likely it is that a store becomes vacant. So, clustering of stores could indeed be successful. However, some other factors could have an influence on this effect that are not considered in this research, because of the unavailability of data. For example, it may be that the retail mix plays a role in predicting this effect. As scientific literature stated that particular store categories cluster together while other types prefer to locate further away from competitors. Besides, it could be stated that retail properties with a known energy label have less chance of becoming vacant compared to retail properties without having a known energy label at all. This gives valuable insights for investors and owners of retail properties. However, more research is needed to give clear distinction effects between the different labels. The analysis over time seems extremely important in explaining the drivers of retail vacancy. Almost all effects are greater in magnitude which is logical since it is measured over 8 years. In addition, the second hypothesis for urbanity level 2 could be supported in the analysis over time and in Model 5 with the simple logit regression with urbanity level 2 as the main effect. However, still essential factors are missing in both the models to explain retail vacancy more properly. Furthermore, it seems questionable whether the results are still relevant post-Covid19. Unless it will be utilized as a comparison with follow-up research on effects during Covid-19.

5.2 Discussion

The results are partly inconsistent with the literature discussed. The inability to reject most hypotheses has some explicable causes. The limitations' discussion starts with the fact that retail vacancy is measured as a simplified binary variable. The length of the vacancy period is thus disregarded. However, the logistic regression approach adopted in this research proved very well. This approach was used to avoid bias from the unavailability of information about the duration length of the vacancy propositions in the data. This way all retail properties that were both vacant and occupied could be included in the analysis. Another limitation regarding the retail stock included in the dataset is that only retail properties that are located in a residence with an inner-city or main shopping area are included in the dataset due to data transmission limits of research organization Locatus. So, not all retail properties in the Netherlands could be included in the dataset. For further limitations regarding this research, a distinction is made between the discussion of the data and methodology.

Regarding the data, having very little information on the retail unit level and observing a lot of missing values in the dataset brings a lot of difficulties explaining retail vacancy. Unfortunately, downsizing the dataset by deleting the missing values may cause a false representation of the population. The availability of particular important control variables such as e-commerce and income were extremely low. Available e-commerce numbers pertain to the whole of the Netherlands and were not even available for all years. On the desired retail unit level, no e-commerce numbers were available at all. So, this extremely important factor could not be included in the model. For other crucial factors little or no data were available on the retail unit level in the dataset, this is a major limitation for this research. Such crucial factors were data on the retail mix, accessibility, and store-image attributes, which provide information on the retail unit level. From analyzing academic literature, it could be concluded that these characteristics may have an influence on retail vacancy. Including all these aspects and moreover preferably not on a lower aggregated level than on retail unit level, could have ensured that the model would be improved in explaining retail vacancy more appropriately.

Besides, assumptions are made regarding the energy labels that have a provisional label. There exists an inaccuracy in it compared to the registered labels. Another inaccuracy regarding energy labels is that no registration year is known for the registered labels of the previous labels of the retail properties. So, an assumption has been made for this study that the previous label for all properties is the same every year until the registration of the current label. Both the use of the provisional labels and the assumption made on the previously registered labels may result in false validity.

Regarding the methodology of the quantitative analysis, an important aspect should be highlighted. In the first instance, using a simple logit model in this research ignores the panel structure of the data. So, changes in occupancy and vacancy of a retail property could not be taken into account. And so, after vacancy was observed, the subsequent change needed to be deleted from the dataset. This is a waste of the data that is available and gives an oversimplified analysis outcome. Therefore afterwards, the panel structure is indeed taken into account by executing a panel logistic regression.

5.3 Policy and research recommendations

Overall, understanding the causes of retail vacancy (thanks to scientific literature and partly thanks to this research) gives valuable insights for the decision of where to locate a particular store and how to decrease the risk of vacancy to the minimum. This is most relevant for investors, owner-occupiers, and as well for tenants. Indirectly it is relevant for municipalities as well, since particular locations could turn out to be unattractive for retailers to locate, however attractive for other functions as houses or offices for example. When these locations have a retail function in the first place the destination plan has to be transformed by the municipality in order to change the function of the property. If this could be detected and transformed in an early phase every party experience benefits from this. Policy recommendations regarding the specific tested hypotheses seem difficult since most hypotheses could not be rejected. Regarding the last hypothesis, an attempt has been made to give policy relevance. It is recommended to actively promote the implementation of an energy label, since from the empirical analysis it could be concluded that retail properties having an energy label have less chance of becoming vacant compared to properties without an energy label known. The advantages of a more energy efficient building are most relevant for owner-occupiers and tenants of a property.

Appropriate follow-up research that appears from academic- and empirical analysis could be on investigating whether stores that sell frequently purchased goods and are easily accessible, lower the odds of becoming vacant. As the empirical analysis showed that the number of stores in a postcode4 area has a slightly negative effect on retail vacancy when considering the panel structure. Additionally, by the fact that this could be explained by academic literature saying that as the consumer purchase frequency and density of customers rises, the density of stores increases as well. However, it remains unclear whether these differences in purchase frequency and accessibility lower the odds of a store becoming vacant. Furthermore, the influence of store-image attributes could be analyzed to gain insights into whether improving store characteristics such as atmospherics, assortments, quantity, and quality lead to higher store profits and consequently to lower retail vacancy. As it could be concluded that a retail property having an energy label lowers the odds of becoming vacant, probably other store characteristics that provide individual store information and also improves individual store characteristics could have an influence on retail vacancy as well. To get an even more accurate picture of store vacancy drivers over time, duration analysis could be applied (van den Berg, 2001). This could be used with logit models to observe and model the properties with multiple changes between occupancy and vacancy over time. Unfortunately, this analysis method was too complex for this study. Lastly, it could be tested whether incorporating the developments and trends during Covid-19 have an association with retail vacancy. For example, stores have suffered from the period when they were obliged to close their doors. For this research, there were no data yet available of these effects, but it seems feasible for future research. Moreover, this way the differences in effects before and during Covid-19 could be compared and learned from.

By analyzing these additional factors, retail vacancy could be explained even more appropriately than it is explained now. By understanding all crucial drivers of retail vacancy, more thoroughly well-considered decisions could be made where to locate a particular store.

5.4 Critical reflection

First, it was of crucial importance to consider how best to measure retail vacancy. It was challenging defining vacancy in the correct matter, since real long-term vacancy is not present in times of transitioning a building to another sector for example. Thus, the exact demarcation when to call a property vacant seemed indefinite. Besides, the data received from research organization Locatus lacks sufficient years for the breakdown by type of vacancy. Due to these reasons, the variable retail vacancy is simplified to a binary variable. And so, the length of the vacancy period is not taken into account in this research.

A great deal of scientific literature turned out to be available regarding the central concepts under review in this research. The subject has been addressed from multiple disciplines in the academic literature. It was challenging to succinctly write down the findings in literature.

In the following phase, collecting data at research organization Locatus was difficult. From the beginning, they were not willing to send all retail stock of the Netherlands, because of the size and comprehensiveness of these data. After one visit, it was not allowed to visit the office anymore due to the Covid-19 measures. As a result, this process took longer than estimated and caused prolongation. In the end, with the help of my supervisor, they were convinced to send the vacant retail stock located in all residences with an inner-city or main shopping area. The occupied stock was obtained through the Geodienst of the University of Groningen and the explanatory variables were collected at the CBS. From all these different sources a complex dataset has been put together and prepared for analysis, which was very time-consuming.

Afterwards, an occupied period came with increasing study and board obligations, so that not all attention could go to writing this thesis. Then, while working on the analysis some difficulties arose due to lack of knowledge about logistic regressions with panel data. After reading several studies with the same sort of data and models it was manageable to include it in this study. However, the unmanageable data has created many difficulties in interpreting the results. After working very intensively on this research in the last recent months, very satisfied it has been handed in.

REFERENCES

Benjamin, J.D., Jud, G.D. and Winkler, D.T., 1998. A Simultaneous Model and Empirical Test of the Demand and Supply of Retail Space. *The Journal of Real Estate Research*, 16(1), pp.1-14.

Benjamin, J.D., Jud, G.D. and Winkler, D.T., 2000. Retail vacancy rates: the influence of national and local economic conditions. *Journal of real estate portfolio management*, 6(3), pp.249-258.

Berben, J. and Kuijpers-van Gaalen, I., 2014. Onderbouwingsmethode definitief energielabel. *BuildDesk.* 11, pp. 1-25

Berg, G.J. van den, 2001. Duration models: specification, identification and multiple durations. *Handbook of econometrics, Elsevier.* 5, pp. 3381-3460.

Bouwinvest, 2020. Dutch Real Estate Market Outlook 2020-2022. *Bouwinvest Real Estate Investors B.V.*

Brouwer, A. and Tool, B., 2018. Eenzijdig aanbod van winkels leidt tot meer winkelleegstand. *Real Estate Research Quarterly.* 17(2), pp. 12-20.

Brubaker, B.T., 2004. Site selection criteria in community shopping centers: implications for real estate developers. (*Doctoral dissertation, Massachusetts Institute of Technology*).

Burt, S. and Sparks, L., 2003. E-commerce and the retail process: a review. *Journal of Retailing and Consumer Services*, 10(5), pp.275-286.

Buursink, J., 1996. The Relative Attractiveness of City Centres as Shopping Locations. *The Preservation of the City Centre as Main Shopping Location*, pp.33-48.

Carroll, G.R., 1985. Concentration and specialization: Dynamics of niche width in populations of organizations. *American journal of sociology*, 90(6), pp.1262-1283.

CBRE, 2020. Real estate market outlook 2020 The Netherlands. *CBRE Research*

CBS, 2019a. Meeste oppervlakteleegstand bij kantoren en winkels. *Centraal Bureau voor Statistiek 2020*

CBS, 2019b. Forecast: 3.5 million single households in 2030. *Centraal Bureau voor Statistiek 2020*

CBS, 2020a. Kerncijfers per postcode. *Centraal Bureau voor Statistiek 2020.* Available at: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode

CBS, 2020b. Consumentenvertrouwen, economisch klimaat en koopbereidheid; gecorrigeerd. *Centraal Bureau voor Statistiek 2020.* Available at: https://opendata.cbs.nl/#/CBS/nl/dataset/83693NED/table?ts=1610027483604

CBS, 2020c. Detailhandel; omzetontwikkeling internetverkopen, index 2015=100. *Centraal Bureau voor Statistiek 2020.* Available at: https://opendata.cbs.nl/#/CBS/nl/dataset/83867NED/table?ts=1610294822594

CBS, 2020d. Regionale kerncijfers Nederland. *Centraal Bureau voor Statistiek 2020.* Available at: https://opendata.cbs.nl/#/CBS/nl/dataset/70072ned/table?ts=1607508462121

CBS, 2021. Consumenten Conjunctuuronderzoek (CCO). *Centraal Bureau voor Statistiek 2021.* Available at: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/consumenten-conjunctuuronderzoek--cco--

Damian, D.S., Curto, J.D. and Pinto, J.C., 2011. The impact of anchor stores on the performance of shopping centres: the case of Sonae Sierra. *International Journal of Retail & Distribution Management.*

DeMaris, A., 1995. A tutorial in logistic regression. *Journal of Marriage and the Family*, pp. 956-968.

Dynamis, 2019. Sprekende cijfers winkelmarkten 2019. *Utrecht: Dynamis B.V. Research*

Evers, D., Tennekes, J. and van Dongen, F., 2014. De bestendige binnenstad. *Den Haag, Planbureau voor de Leefomgeving*

Fuerst, F. and McAllister, P., 2011. The impact of Energy Performance Certificates on the rental and capital values of commercial property assets. *Energy policy*, 39(10), pp.6608-6614.

Geltner, D., Miller, N.G., Clayton, J. and Eichholtz, P., 2001. Commercial real estate analysis and investments (Vol. 1). *Cincinnati, OH: South-western.*

Genecon, L.L.P., and Partners, 2011. Understanding high street performance. *Department for Business, Innovation and Skills*

Ghosh, A., 1986. The value of a mall and other insights from a revised central place model. *Journal of retailing*, 62(1), pp.79-97.

Gould, E.D., Pashigian, B.P. and Prendergast, C.J., 2005. Contracts, externalities, and incentives in shopping malls. *Review of Economics and Statistics*, 87(3), pp.411-422.

Haar, F. ter, and Quix, F., 2020. Retail postcorona – impactanalyse. *Retailland*

Hill, R.C., Griffiths, W.E. and Lim, G.C., 2018. Principles of econometrics. *John Wiley & Sons*

Hollander, J.B., Hartt, M.D., Wiley, A. and Vavra, S., 2018. Vacancy in shrinking downtowns: a comparative study of Québec, Ontario, and New England. *Journal of Housing and the Built Environment*, 33(4), pp.591-613.

Hsiao, C., 1996. Logit and probit models. *The Econometrics of Panel Data*, pp. 410-428. *Springer*

Huisman, J., 2020. Binnen twee jaar verdwijnt een kwart van de non-food winkels. *Vastgoedmarkt.* Available at: *https://www.vastgoedmarkt.nl/financieel/nieuws/2020/07/binnen-twee-jaar-verdwijnt-een-kwart-van-de-non-food-winkels-101155534*

Hurd, M. and Rohwedder, S., 2003. The retirement-consumption puzzle: Anticipated and actual declines in spending at retirement. *NBER Working Paper 9586*.

Karande, K. and Lombard, J.R., 2005. Location strategies of broad-line retailers: an empirical investigation. *Journal of Business Research*, 58(5), pp.687-695.

Kok, N. and Jennen, M., 2012. The impact of energy labels and accessibility on office rents. *Energy Policy*, 46, pp.489-497.

Kumar, V. and Karande, K., 2000. The effect of retail store environment on retailer performance. *Journal of business research*, 49(2), pp.167-181.

KvK, 2020. Standaard Bedrijfsindeling 2008 – update 2019. *Kamer van Koophandel.* Available at: https://www.kvktoegankelijk.nl/sbi2019nederlands/

Leeuwen, N. van, 2019. Statistische gegevens per vierkant en postcode 2018-2017-2016-2015. *Centraal Bureau voor Statistiek 2019*

Liu, C.H., Rosenthal, S.S. and Strange, W.C., 2018. The vertical city: Rent gradients, spatial structure, and agglomeration economies. *Journal of Urban Economics*, 106, pp.101-122.

Locatus, 2017. Beschrijving Verkooppunt Verkenner Nederland. *Woerden, Locatus.* 4, pp. 1-20

Locatus, 2018. Beschrijving Verkooppunt Verkenner. *Woerden, Locatus*. 11, pp. 1-24.

Locatus, 2019. Het winkelverkoopvloeroppervlak detailhandel. *Woerden, Locatus.* 5, pp. 1-2.

Locatus, 2020. Factsheet Retaildata Benelux 2020. *Woerden, Locatus.*

McKibbin, Warwick. J., and Stoeckel, A., 2009. Modelling the global financial crisis. *Oxford Review of Economic Policy*, 25(4), pp. 581-607.

Midi, H., Sarkar, S.K. and Rana, S., 2010. Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics, 13(3), pp.253-267.

Mingardo, G. and van Meerkerk, J., 2012. Is parking supply related to turnover of shopping areas? The case of the Netherlands. *Journal of Retailing and Consumer Services*, 19(2), pp.195-201.

Myers, D. and Wyatt, P., 2004. Rethinking urban capacity: identifying and appraising vacant buildings. *Building Research & Information,* 32(4), pp.285-292.

NRW, 2014. De noodzaak van structurele vernieuwing van de winkelstructuur: vitaliteit en transitie binnen heldere kaders. *Utrecht: Taskforce Dynamische Retailontwikkeling - Nederlandse Raad Winkelcentra.*

Ouwehand, A., 2018. Drie argumenten om leegstand nu aan te pakken. *Vastgoedmarkt.* Available at: *https://www.vastgoedmarkt.nl/beleggingen/nieuws/2018/07/drie-argumenten-om-leegstand-nu-aan-te-pakken-101134771*

Parker, A.J., 1992. Consumer Attitudes to the Retail Environment: A Case Study of the City Centre. *In*: G. Heinritz, ed., *The Attraction of Retail Locations*, ed., pp. 155-172.

Paul, J., Sankaranarayanan, K. G., and Mekoth, N., 2016. Consumer satisfaction in retail stores: Theory and implications. *International Journal of Consumer Studies*, 40(6), pp.635-642.

PBL & ASRE, 2013. Gebiedsontwikkeling en commerciële vastgoedmarkten - Een institutionele analyse van het (over)aanbod van winkels en kantoren. *Den Haag/Amsterdam: Planbureau voor de Leefomgeving/Amsterdam School of Real Estate*

Rhodes, C. and Brien, P., 2014. The retail industry: statistics and policy. *House of Commons Library Briefing Paper*

Rijksoverheid, 2020. EP-Online. Available at: https://www.ep-online.nl/ep-online/PublicData

RPC, 2020. Leegstaande panden door crisis trekken criminaliteit aan. *Regionaal Platform Criminaliteitsbeheersing Noord-Nederland*

RVO, 2018. Welke utiliteitsgebouwen(delen) zijn labelplichtig? *Rijksdienst voor Ondernemend Nederland*

Schmidt, C.G., 1983. Location Decision-Making Within a Retail Corperation. *Journal of Regional Analysis and Policy*, 13(1100-2016-89559), pp.60-71.

Sevtsuk, A., 2014. Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research*, 34(4), pp.374-393.

Slob, G., 2020. Na jaren van herstel neemt de leegstand in Nederland weer flink toe. *Woerden, Locatus*

Stoltzfus, J.C., 2011. Logistic regression: a brief primer. *Academic Emergency Medicine*, 18(10), pp.1099-1104.

Teller, C. and Elms, J., 2010. Managing the attractiveness of evolved and created retail agglomerations formats. *Marketing Intelligence & Planning*

Teulings, C.N., Ossokina, I.V. and Svitak, J., 2018. The urban economics of retail. *The Hague, CPB Netherlands Bureau for Economic Policy Analysis*

Turhan, G., Akalın, M. and Zehir, C., 2013. Literature review on selection criteria of store location based on performance measures. *Procedia-Social and Behavioral Sciences*, 99, pp.391-402.

Velzen, J., 2020. Oplossing voor leegstaande winkelpanden: bouw ze om tot woningen. *Trouw.* Available at: *https://www.trouw.nl/nieuws/oplossing-voor-leegstaande-winkelpanden-bouw-ze-om-tot-woningen~bfef3512/*

Ward, M.R., 2001. Will online shopping compete more with traditional retailing or catalog shopping? *Netnomics*, 3(2), pp.103-117.

Weltevreden, J.W.J., 2006. City centres in the internet age: Exploring the implications of b2c e-commerce for retailing at city centres in the Netherlands. (*Doctoral dissertation, Utrecht University*).

Weltevreden, J.W. and Rietbergen, T.V., 2007. E-Shopping versus City Centre Shopping: The Role of Perceived City Centre Attractiveness. *Tijdschrift voor economische en sociale geografie,* 98(1), pp.68-85.

Wesselink, J.W., 2004. Wonen boven winkels, minder leegstand, meer leefbaarheid. *ROM*, 22, pp.8-10.

Zhang, D., Zhu, P. and Ye, Y., 2016. The effects of E-commerce on the demand for commercial real estate. *Cities*, 51, pp.106-120.

APPENDICES

Appendix A: Operationalization variables

Retail vacancy

Table 3. *Proportion vacant properties for all years in total in panel dataset*

Vacancy (no, yes)	**Frequency**	**Percentage**
0	703,231	90.19
1	76,492	9.81
Total	779,723	100.00

Table 4. *Proportion vacant properties per year in panel dataset*

Vac.	2012		2013		2014		2015		2016	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
0	94,243	92.87	91,212	92.75	87,824	91.91	82,654	90.58	78,118	90.02
1	7,235	7.13	7,127	7.25	7,729	8.09	8,599	9.42	8,663	9.98
Tot.	101,478	100	98,339	100	95,553	100	91,253	100	86,781	100

2017		2018		2019		2020	
Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
73,453	89.73	68,659	89.32	64,958	89.29	62,110	82.99
8,411	10.27	8,209	10.68	7,793	10.71	12,726	17.01
81,864	100	76,868	100	72,751	100	74,836	100

Table 5. *Proportion of vacant properties in each shopping area type*

Shopping area type	**Frequency**	**Percentage**
Central	42,933	56.13
Supporting	19,493	25.48
Other	2,284	2.99
Scattered	11,782	15.40
Total	76,492	100.00

Urbanity

Table 6. *Distribution of urbanity in panel dataset on both postcode4- and municipality level*

			Postcode4 level		*Municipality level*	
Urbanity class	**Indicator**	**Number of addresses/km^2**	**Frequency**	**Percentage**	**Frequency**	**Percentage**
1	Very urban	≥ 2500	284,084	56.25	306,125	40.84
2	Highly urban	1500-2500	145,411	28.79	283,595	37.83
3	Moderately urban	1000-1500	51,180	10.13	111,727	14.90
4	Little urban	500-1000	19,906	3.94	44,629	5.95
5	Non-urban	< 500	4,474	0.89	3,539	0.47
Total			505,055	100.00	749,615	100.00

Building energy efficiency

Table 7. *Energy index linked with energy label and proportion in dataset*

Energy index	**Energy label**	**Frequency**	**Percentage**
≤ 0,5	A++++	57	0.02
≤ 0,5	A+++	134	0.04
≤ 0,5	A++	797	0.27
0,51-0,70	A+	2,368	0.79
0,71-1,05	A	113,856	38.11
1,06-1,15	B	20,848	6.98
1,16-1,30	C	44,131	14.77
1,31-1,45	D	18,106	6.06
1,46-1,60	E	18,844	6.31
1,61-1,75	F	26,970	9.03
≥ 1,75	G	52,613	17.61
Total		298,724	100.00

Economy - Questionnaire

1. In general, do you think that the economic situation of our country has improved, gotten worse or stayed the same during the last twelve months?
2. And what do you think of the next twelve months? In general, will the economic situation in the Netherlands improve, deteriorate or remain the same?
3. Do you consider the financial situation of your household has improved, gotten worse or remained unchanged during the last twelve months?
4. What do you expect from the financial situation of your household? Will it get better, get worse, or remain unchanged during the next twelve months?
5. When it comes to furniture, a washing machine, a television and other durable items? Do you think now is a favorable or unfavorable time for people to make so many large purchases?

Appendix B: Histograms store agglomeration

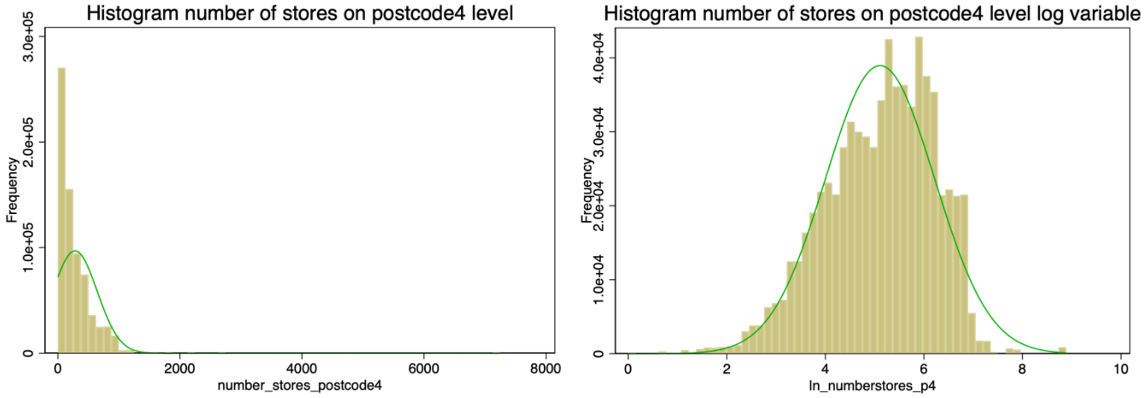

Figure 3. *Histograms number of stores postcode4 level log transformation*

Appendix C: Testing logistic regression assumptions

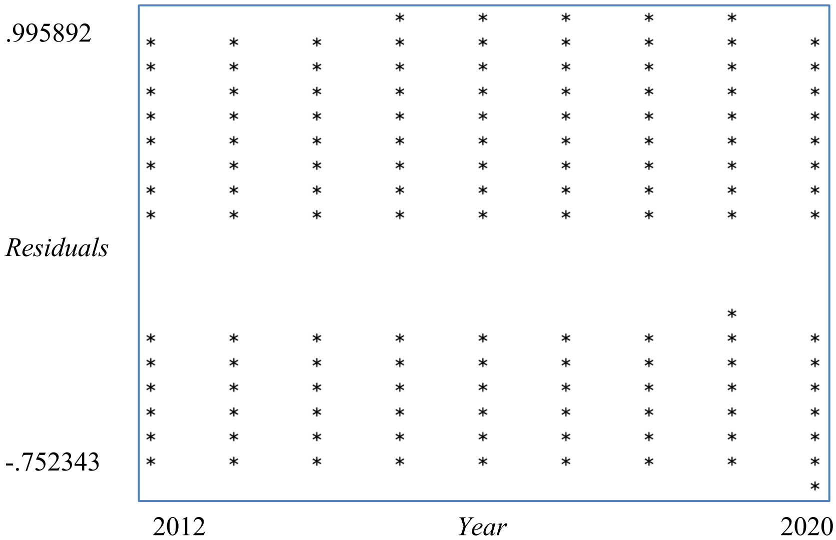

Figure 4. *Residual plot against time*

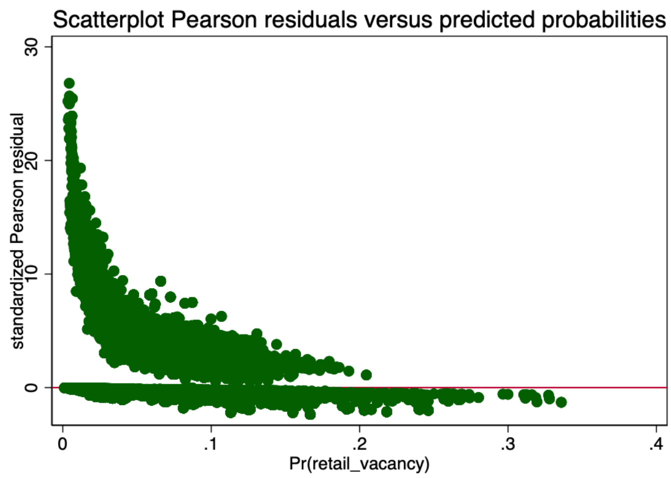

Figure 5. *Scatterplot Pearson residuals*

Table 9. *Box-Tidwell regression model on main independent variables*

Logistic regression	Number of obs = 679,814
	LR chi2(10) = 11463.50
	Prob > chi2 = 0.0000
Log likelihood = -135186.14	Pseudo R2 = 0.0407

retail_vacancy	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Iln_n__1	0.011	0.014	0.820	0.412	-0.016	0.039
Iln_n_p1	-0.000	0.006	-0.040	0.969	-0.013	0.012
Iln_na_1	-0.002	0.020	-0.120	0.904	-0.042	0.038
Iln_nap1	0.000	0.004	0.010	0.992	-0.008	0.008
urbanitylevel						
2	0.201	0.013	15.900	0.000	0.176	0.225
3	0.257	0.016	15.720	0.000	0.225	0.289
4	0.274	0.023	11.800	0.000	0.229	0.320
5	-0.361	0.092	-3.910	0.000	-0.542	-0.180
energy_efficient	-1.174	0.017	-67.740	0.000	-1.208	-1.140
energy_inefficient	-1.487	0.025	-60.440	0.000	-1.535	-1.438
_cons	-2.627	0.011	-248.450	0.000	-2.648	-2.607
ln_numbers~4	.0608955	.0051075	11.92	Nonlin. dev.	8.350	(P = 0.004)
p1	1.742475	.5775187				
ln_numbers~6	.0720692	.0057399	12.56	Nonlin. dev.	431.207	(P = 0.000)
p1	-1.024466	1.538383				

Table 10a. *Collinearity Diagnostics – VIF scores*

Variable	**VIF**	**SQRT VIF**	**Tolerance**	**R-Squared**
ln_numberstores_p4	2.05	1.43	0.4888	0.5112
ln_numberstores_p6	1.17	1.08	0.8526	0.1474
lowenergy	1.14	1.07	0.8808.	0.1192
mediumenergy	1.07	1.04	0.9320	0.0680
highenergy	1.11	1.05	0.9012	0.0988
inh_upto_15years	4.81	2.19	0.2078	0.7922
inh_15to25years	2.78	1.67	0.3592	0.6408
inh_25to45years	6.08	2.46	0.1646	0.8354
inh_45to65years	7.39	2.72	0.1353	0.8647
inh_65yearsandolder	2.72	1.65	0.3679	0.6321
populationdensity_n_inh	3.41	1.85	0.2932	0.7068
populationgrowth_relative	1.92	1.39	0.5205	0.4795
privatehouseholds_n	3.81	1.95	0.2627	0.7373
householdsize_avg	3.88	1.97	0.2574	0.7426
economicsituation_next12m	6.16	2.48	0.1623	0.8377
urbanitylevel	3.24	1.80	0.3088	0.6912
highincome	1.77	1.33	0.5643	0.4357
internetturnover_index	3.94	1.98	0.2538	0.7462
turnoverdev_toyearbefore	6.49	2.55	0.1542	0.8458
lnfloorspace	1.13	1.06	0.8828	0.1172
Mean VIF	3.30			

Table 10b. *Collinearity Diagnostics final model*

Variable	**VIF**	**SQRT VIF**	**Tolerance**	**R-Squared**

ln_numberstores_p4	1.96	1.40	0.5094	0.4906
ln_numberstores_p6	1.16	1.08	0.8618	0.1382
energy_efficient	1.10	1.05	0.9124	0.0876
energy_inefficient	1.09	1.04	0.9174	0.0826
inh_upto_15years	4.68	2.16	0.2134	0.7866
inh_15to25years	2.81	1.68	0.3564	0.6436
inh_25to45years	6.00	2.45	0.1668	0.8332
inh_45to65years	7.43	2.73	0.1345	0.8655
inh_65yearsandolder	2.66	1.63	0.3753	0.6247
populationdensity_n_inh	3.29	1.81	0.3037	0.6963
privatehouseholds_n	3.37	1.84	0.2964	0.7036
householdsize_avg	3.05	1.75	0.3275	0.6725
economicsituation_next12m	1.09	1.04	0.9201	0.0799
financialsituation_last12m	1.11	1.05	0.9027	0.0973
urbanitylevel	3.07	1.75	0.3253	0.6747
lnfloorspace	1.11	1.06	0.8981	0.1019
Mean VIF	2.81			

Appendix D: Regression results

Table 11. *Linktest results Model 1 simple logit regression*

Logistic regression	Number of obs	= 668,583
	LR chi2(2)	= 17462.01
	Prob > chi2	= 0.0000
Log likelihood = -121163.06	Pseudo R2	= 0.0672

retail_vacancy	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
_hat	0.699	0.055	12.64	0.000	0.591	0.807
_hatsq	-0.050	0.009	-5.50	0.000	-0.067	-0.032
_cons	-0.428	0.081	-5.27	0.000	-0.587	-0.268

Table 13. *Logistic regression (postcode4 level)*

	Model 1		Model 2	
retail_vacancy	Coef.	St.Err.	Coef.	St.Err.
ln_numberstores_p4	$3.533\text{x}10^{-4}$	.009		
ln_numberstores_p6			.085***	.007
energy_efficient	-1.433***	.022	-1.465***	.022
energy_inefficient	-1.809***	.031	-1.795***	.031
inh_upto_15years	$3.44\text{x}10^{-5}$	$2.15\text{x}10^{-5}$	$4.29\text{x}10^{-5}$**	$1.86\text{x}10^{-5}$
inh_15to25years	$1.141\text{x}10^{-4}$***	$1.2\text{x}10^{-5}$	$1.132\text{x}10^{-4}$***	$1.21\text{x}10^{-5}$
inh_25to45years	$6.78\text{x}10^{-5}$***	$1.01\text{x}10^{-5}$	$6.64\text{x}10^{-5}$***	$9.34\text{x}10^{-6}$
inh_45to65years	$-2.439\text{x}10^{-4}$***	$1.99\text{x}10^{-5}$	$-2.359\text{x}10^{-4}$***	$1.97\text{x}10^{-5}$
inh_65yearsandolder	$2.336\text{x}10^{-4}$***	$1.53\text{x}10^{-5}$	$2.12\text{x}10^{-4}$***	$1.5\text{x}10^{-5}$
populationdensity_~h	$-2.405\text{x}10^{-4}$***	$5.18\text{x}10^{-6}$	$-2.402\text{x}10^{-4}$***	$5.18\text{x}10^{-6}$
economicsituat~xt12m	.002	.003	.004	.003
financialsitua~st12m	-.067***	.011	-.071***	.011
lnfloorspace	-.271***	.006	-.266***	.006
year				
2013	-8.43***	.078	-.876***	.078
2014	-1.49***	.365	-1.66***	.364
2015	-.572***	.188	-.664***	.187
2016	-.151	.118	-.213*	.118
2017	.286***	.079	.24***	.079

2018	.482***	.063	.447***	.063
2019 (omitted)				
2020 (omitted)				
constant	-1.407***	.077	-1.564***	.065
number of obs.	584085		584085	
chi2	12654.987		12813.172	
prob > chi2	0.0		0.000	
pseudo R2	0.086		0.087	

*Note: The dependent variable is the binary variable retail_vacancy, indicating whether a retail property is vacant or not vacant. *, **, *** are significant at 10%, 5% and 1% respectively. The standard error are clustered on ID.*

Table 14. *Logistic regression with three distinction groups of building energy efficiency (municipality level)*

retail_vacancy	Coef.	St.Err.
ln_numberstores_p4	.031***	.008
lowenergy	-1.239***	.024
mediumenergy	-1.23***	.028
highenergy	-1.885***	.034
inh_upto_15years	3.3×10^{-5}	2.09×10^{-5}
inh_15to25years	-1.92×10^{-5}	1.33×10^{-5}
inh_25to45years	5.36×10^{-5}***	1.06×10^{-5}
inh_45to65years	-1.656×10^{-4}***	1.79×10^{-5}
inh_65yearsandolder	1.941×10^{-4}***	1.38×10^{-5}
populationdensity_~h	-4.74×10^{-5}***	7.01×10^{-6}
privatehouseholds_n	-2.21×10^{-6}***	9.03×10^{-8}
householdsize_avg	.012	.057
economicsituat~xt12m	-.015***	.002
financialsitua~st12m	.024**	.01
urbanitylevel		
2.urbanitylevel	-.192***	.024
3.urbanitylevel	-.262***	.033
4.urbanitylevel	-.244***	.041
5.urbanitylevel	-.99***	.108
lnfloorspace	-.29***	.005
year		
2013	-.332***	.069
2014	1.081***	.322
2015	.656***	.166
2016	.509***	.106
2017	.46***	.072
2018	.574***	.058
2019 (omitted)		
2020 (omitted)		
constant	-.97***	.131
number of obs.	668583	
chi2	12432.865	
prob > chi2	0.000	
pseudo R2	0.068	

*Note: The dependent variable is the binary variable retail_vacancy, indicating whether a retail property is vacant or not vacant. *, **, *** are significant at 10%, 5% and 1% respectively. The standard errors are clustered on ID.*

Table 15. *Logistic regression for robustness analysis different shopping area types(municipality level)*

retail_vacancy	Coef.	St.Err.	Coef.	St.Err.
ln_numberstores_p6	.225***	.011	.225***	.011
energy_efficient	-1.417***	.036	-1.417***	.036
energy_inefficient	-2.005***	.05	-2.005***	.05
inh_upto_15years	1.665×10^{-4}***	2.63×10^{-5}	1.665×10^{-5}***	2.63×10^{-5}
inh_15to25years	9.52×10^{-5}***	2.22×10^{-5}	9.52×10^{-5}***	2.22×10^{-5}
inh_25to45years	3.4×10^{-5}**	1.34×10^{-5}	3.4×10^{-5}**	1.34×10^{-5}
inh_45to65years	-2.059×10^{-4}***	2.78×10^{-5}	2.059×10^{-4}***	2.78×10^{-5}
inh_65yearsandolder	4.13×10^{-5}*	2.45×10^{-5}	4.13×10^{-5}*	2.45×10^{-5}
populationdensity_~h	-4.76×10^{-5}***	1.06×10^{-5}	-4.76×10^{-5}***	1.06×10^{-5}
privatehouseholds_n	-1.77×10^{-6}***	1.66×10^{-7}	-1.77×10^{-6}***	1.66×10^{-7}
householdsize_avg	1.176***	.135	1.176***	.135
economicsituat~xt12m	-.031***	.004	-.031***	.004
financialsitua~st12m	.093***	.016	.093***	.016
urbanitylevel				.
2.urbanitylevel	-.114**	.048	-.114**	.048
3.urbanitylevel	-.803***	.104	-.803***	.104
4.urbanitylevel	-1.068***	.192	-1.068***	.192
5o.urbanitylevel				.
lnfloorspace	-.088***	.008	-.088***	.008
year				.
2013.year	.118	.113	.118	.113
2014.year	3.385***	.524	3.385***	.524
2015.year	1.81***	.273	1.81***	.273
2016.year	1.185***	.175	1.185***	.175
2017.year	.753***	.124	.753***	.124
2018.year	.87***	.101	.87***	.101
2019o.year				.
2020o.year				.
area				
1.area	-.074	.15		.
2.area	-.23	.141	-.156***	.048
3.area			.074	.15
Constant	-3.874***	.316	-3.948***	.295

*Note: The dependent variable is the binary variable retail_vacancy, indicating whether a retail property is vacant or not vacant. *, **, *** are significant at 10%, 5% and 1% respectively. The standard errors are clustered on ID.*