The relationship between attractiveness of retail areas and their performance

July 18th, 2022

N. H. Hoomans

Master Thesis Msc. Real Estate Studies

COLOFON

Title	The relationship between attractiveness of retail areas and their performance
Version	Final
Author	N. H. (Nynke) Hoomans
Student number	S4574028
Supervisor	Frans Schilder
Assessor	Ed Nozeman
E-mail	n.h.hoomans@student.rug.nl
Date	July 18th, 2022

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'

ABSTRACT

Inner cities are subject to the changes occurring in society. The deterioration that occurs when vacancy increases may indicate a retail area's decreasing attractiveness. However, retail areas are performing above prior expectations. It is essential to understand the impact of changing retail circumstances and shocks like Covid-19 on the performance of retail areas. According to Reilly's law of retail gravitation, attractiveness is determined by only size and distance of the retail area. To investigate this phenomenon, the relationship between attractiveness indicators of a retail area and its performance is researched. Using a multiple linear regression method, the results show that size and distance explain 53 percent of footfall and 68 percent of turnover, indicating that Reilly's law cannot entirely explain attractiveness and other variables should be incorporated. Location effects add a significant amount to the explanatory power. Furthermore, retail areas are sensitive to the size of the municipality they are located in. A small retail area in a large municipality generates a higher turnover than a small retail area in a small municipality. A recommendation for further research is to include a longer time frame to compare reactions to economic and social circumstances.

Keywords: Retail, attractiveness, footfall, gravitation theory

CONTENTS

1. INTRODUCTION

1.1. Motivation

Inner cities are subject to the changes occurring in society. Retail real estate has been coping with rising vacancy for the past years. It is expected that vacancy rates will continue to grow. Some research even expected an increase to forty percent in 2022 (NOS, 2020). However, in 2021 vacancy rates decreased from 7.5 percent at the beginning of the year to 6.8 percent by the end of the year. This decline can partly be explained by nearly 2000 properties being extracted from the market. In April 2022, 6.4 percent of Dutch retail space was vacant (Locatus, 2022). Furthermore, retail real estate values have recently been recovering. After substantial decreases in value in the past few years, the depreciation of shopping center values came to a stop (Vastgoedmarkt, 2022). Additionally, customer spending has been increasing, even though consumer confidence is at an all-time low (NOS, 2022b; CBS, 2022). This difference from the expectation shows that predictions for retail real estate performance are difficult to establish.

Many factors are related to the changing performance of retail areas. Examples include the rise of e-commerce, the 2008 economic crisis, demographic changes, and changes in consumer demand (MDBS, 2014). The factors mentioned above are known by many at this point and can be labelled as long-term factors. More recently, Covid-19 has been indicated as a cause for lower turnover, lower footfall, and higher vacancy rates. As a result of the pandemic, many stores have seen sales drop and even stores going bankrupt (Talen & Park, 2021). City centers become less attractive when many units are left vacant because of deterioration (Zhang et al., 2019). This may suggest that the increasing vacancy rates lead to a further decrease in demand because of this unattractiveness, which will in turn lead to a decrease in footfall and turnover.

On the other hand, outlooks for retail real estate performance are positive. Footfall in retail areas is not yet at the same level as pre-Covid-19, but the numbers are recovering faster than expected. Furthermore, customers spend more per person (KSO, 2022). Retailers in Dutch shopping streets indicated that customers are returning to the stores, and in some stores turnovers are higher than pre-Covid-19 (NOS, 2022b). When walking through a shopping area, it is visible that the retail environment is adjusting to changes in demand, and as a result, the composition of segments has changed. For example, the convenience market is gaining ground, and the number of convenience stores like bakeries, supermarkets and liquor stores has increased (NOS, 2022a). Retail areas are adjusting to the changing retail climate. Repositioning retail real estate is necessary in a highly dynamic market. It is needed to adapt to the changing preferences of users. Repositioning is changing the product, which is the shopping area, to meet the users' preferences and expectations (Nanda et al., 2021).

Overall, retail areas seem to have resurrected against all prior expectations. Customers still value physical engagement, the ability to test products in a physical store and to gather advice before making a purchase (Zhang et al., 2022). The difference between prior expectations and the current state of retail real estate evokes the question how attractiveness of physical retail real estate influences performance. Therefore, this research focuses on gaining more knowledge on the retail areas of today by determining if there is a relationship between the performance of retail areas and the attractiveness of retail real estate in the Netherlands.

1.2 Academic relevance

Much research has been done on the attractiveness of retail real estate. Footfall data, turnover and vacancy rates are often used to indicate the economic performance of retail real estate. In 2019, research was published on the effect of increasing footfall on vacancy rates on the national level for the Netherlands. It was found that when footfall increases, vacancy rates will decrease. However, the increase in footfall is higher than the decrease in vacancy. The authors conclude that shops tend to cluster together to benefit from positive externalities, like higher footfall and a lower chance of vacancy (Koster, Pasidis, van Ommeren, 2019). The benefits of inter-store externalities for shops have been widely acknowledged in research. For example, in 1986, Gabszewicz and Thisse found that store clustering provides positive externalities (Gabszewicz & Thisse, 1986). Furthermore, in research by Meija and Eppli (2003), it is found that when shops cluster together, this positively impacts in-store sales (Meija & Eppli, 2003).

From the previous part, it becomes clear that shops cluster in specific places to make use of the positive externalities like high footfall because of other shops located at the same place. This statement is easily relatable to core shopping areas, but how does this relate to other locations? Recent research on the UK retail market finds that vacancy rates are higher in 'micro-locations', which are locations characterized as smaller, secondary centers of a larger area. Property owners in these areas experience more difficulty finding tenants. These micro-locations also have a different composition of types of shops than core areas (Philp, Dolega and Green, 2021). Also, Mumford et al. (2017) found that footfall is relatively higher in large city and town centers than in smaller areas (Mumford et al., 2017).

Research by Koster et al. from 2021 focused on the effect of Covid-19 policies on retail real estate using footfall data. This paper found that footfall decreased significantly due to Covid-19 policies (Koster et al., 2021). The same conclusion was formed from research on high streets in the UK, where footfall fell by 57% - 75% during lockdown measures. An interesting outcome from this study is that smaller centers were affected less by the measures, which can possibly be related to smaller centers serving a lower-distance audience (Enoch et al., 2021). As mentioned above, visitors of retail areas in the Netherlands tend to shop closer to home since the Covid-19 crisis. This finding is confirmed by market research, which also found that neighborhood centers suffered the least from Covid-19 measures, especially neighborhood convenience centers. Fashion stores experienced the most significant decline due to both Covid-19 and the increase of e-commerce (BouwInvest, 2020).

An essential indicator of a market's performance is investments. Investors have focused on shopping centers serving a local audience and core shopping streets. Overall, big cities are expected to possess the best performing retail real estate when a higher share of retail turnover will be generated in e-commerce and the demand for retail space decreases (BouwInvest, 2021). However, investments in high streets have decreased since 2017, when there was a peak in high street investment of over two billion euros. In 2021, the most significant part of retail investment was in neighborhood centers, and total retail investment was only 1.6 billion euros. On the other hand, because of the increasing demand for e-commerce, investment in[1] logistics real estate has increased significantly in 2021 (Cushman & Wakefield, 2022; CBRE, 2022).

Customer preferences are one of the most significant changes in the current retail environment. Nevertheless, the physical store remains an important factor in retailers' business models. Customers still value the physical store, while online shopping has been possible for years already (Nanda et al., 2021), showing that the physical store keeps a level of attractiveness. Attractiveness is thus an essential variable in explaining the changing retail real estate sector and its performance.

From this concise literature review, it becomes clear that much research has been done on both the mechanisms behind the performance and attractiveness of retail areas. However, no research has been done on the relationship between attractiveness, footfall and turnover for the retail market in the Netherlands[1]. For retail areas to function to their best potential, it is vital to understand how attractiveness and performance relate. This research aims to explain the mechanisms behind retail real estate by finding the relationship between retail areas' attractiveness and performance.

1.3. Research problem statement

The research aim of this study is to determine if the performance of retail real estate is influenced by the attractiveness of the shopping area. Following from this, the research aims to contribute to the knowledge on this subject and better understand retail centers to make sustainable future planning decisions. To achieve this, the following research question was established:

[1] Based on results found on Google Scholar using the keywords attractiveness, footfall, vacancy, turnover, performance combined with retail, retail areas, shopping center, shopping area, COVID-19, shopping, stores, cities.

To what extent does the attractiveness of retail real estate in the Netherlands influence the performance of retail areas?

To answer the research question, three sub-questions were formulated:

1. What are the determinants of the attractiveness of retail areas?

   This question functions to gain more understanding of the mechanisms behind attractiveness. The method for answering the question will be literature research. In the literature review, some of the determinants have already been mentioned. This part will go deeper into the subject and will also function to come up with reliable control variables for the following two research questions. Research on the determinants of the attractiveness of shopping areas has already been done, so this previous research will be used to answer the question. As a result, a clear overview of attractivity determinants will be established to support the following two sub-questions.

2. What is the relationship between attractiveness and performance of retail areas in the Netherlands?

   A multiple linear regression analysis will be executed to answer the second question. Data are needed on vacancy rates, footfall, turnover, and the variables that explain attractiveness. Footfall is included since the goal of creating an attractive retail area is to attract customers, so to be able to say something about attractiveness, it is essential to include footfall. Turnover is a measure of the retail areas' profitability. This data is not publicly available but will be provided via an internship placement at Colliers in Amsterdam. Furthermore, control variables found in sub-question one will be added to ensure a valid outcome. Further explanation of the data and method will be provided in chapter three.

3. How does this relationship differ between municipalities and types of retail areas?

   The outcomes of the previous research questions are simply an average of all the locations entered in the model. However, there may be significant differences between areas. It is interesting to understand the differences between areas since the outcome may not represent all areas and similarly for all types of retail areas. The result from research question two will describe the average relationship between performance and the other variables added to the model. It will be fascinating to see if the relationship varies for these different types of retail areas. The outcome of question two is needed and so are the same data as a pooled sample to answer this question.

Figure 1.1 shows the conceptual model for this research. Starting from the left side - distance, vacancy, size and customer valuations are used to measure attractiveness. These variables were chosen as the outcome of the first research question; what are the determinants for the attractiveness of retail areas? From this, research question two can be answered; what is the relationship between the attractiveness of retail areas and their performance? Next, location and type of retail area are included as control variables and also used in the final research question; how does the relationship differ between areas and type of retail area?

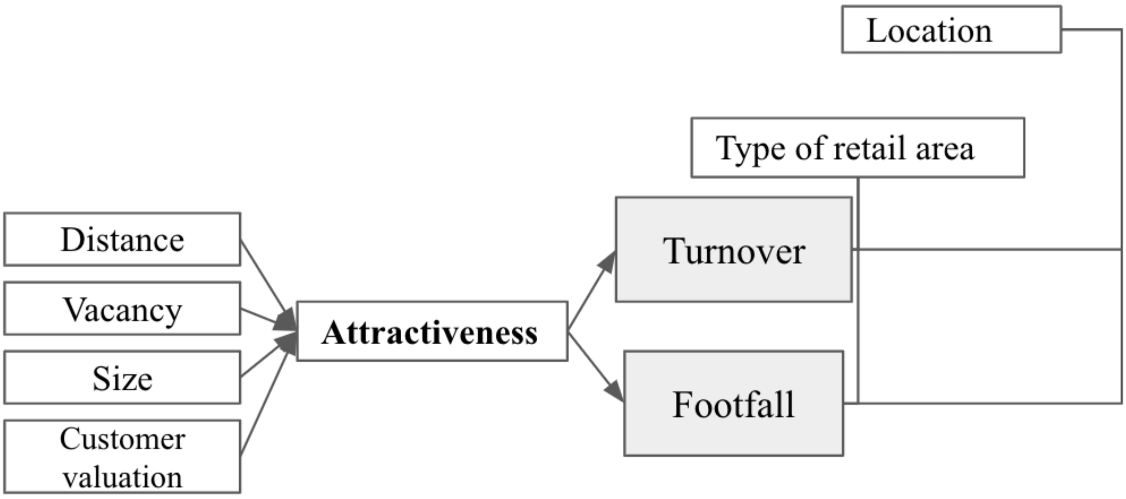

Figure 1.1*:* Conceptual model

The thesis is structured as follows: Relevant theories and literature are discussed in the next chapter. After that, chapter 3 contains an explanation of the data and methods used to conduct this research. Results are presented in chapter 4 and will be discussed in chapter 5. Finally, the conclusions are discussed in chapter 6.

2. THEORY, LITERATURE REVIEW AND HYPOTHESES

2.1 Retail gravitation theory

Retail gravitation theory is widely used by scholars researching the performance of retail areas. Many researchers have touched on the subject of retail gravitation. Overall, Reilly can be seen as the founder of retail gravitation theory. Reilly's law of retail gravitation suggests that customers are willing to travel a further distance to go to a larger shopping area. The size of the shopping center or area is an indicator of attractiveness. The assumptions in Reilly's model are that there are no barriers when travelling to a location so that the latitudinal distance can be used, and the customer is indifferent in its choice between cities (Reilly, 1931).

Reilly (1931) finds that "retail business gravitates from smaller cities and towns to larger cities in accordance with a definite law (pp. 5)". The breaking point of trade between a larger city and a smaller city is the distance from where customers will go to which city. Logically, the breaking point is further from the larger city than the smaller city. From this expression follows Reilly's law of retail gravitation:
"Two cities attract retail trade from any intermediate city or town in the vicinity of the breaking point, approximately in direct proportion to the populations of the two cities and in inverse proportion to the square of the distances from these two cities to the intermediate town (Reilly, 1931, pp. 9)"

So, according to Reilly's law of retail gravitation, distance and size of the city or retail area are the variables explaining attractiveness. It may seem illogical that just distance and size can explain the attractiveness of a city. However, Reilly's findings show that distance and size are the primary drivers of retail gravitation, since the inconvenience, time and cost it takes to travel to another city are all dependent on distance, and availability of retail-mix is dependent on size (Reilly, 1931).

Reilly's usage of the word "law" shows that he saw his work as an undeniable fact, which is also mentioned in his book. However, that publication contains a statement of being open to new findings and additions to the law. Since Reilly, researchers from different periods have tested the law of retail gravitation and added adjustments.

First, Converse (1949) added to retail gravitation theory by slightly adjusting the original theory. The first step in Converse's theory was to determine the boundaries of trade for the retail area, which can be established using distance and population. From this, Converse's New Law of Retail Gravitation follows:
"a trading center and a town in or near its trade area divide the trade of the town approximately in direct proportion to the populations of the two towns and inversely as the squares of the distance factors, using four as the distance factor of the home town" (Converse, 1949, p. 382).

The usage of 4 as the distance factor of the hometown was determined because Converse found that the outcome for distance in Reilly's original formula tended to be close to four. The adjusted theory by Converse differs from Reilly's theory, but Converse also uses just distance and size as the variables explaining the attractivity of a town.

In 1964, Huff published another retail gravitation theory known as "Huff's model of trade area attraction". Huff notes that Converse's formula for the boundary of a trading area does not hold when boundaries for different areas overlap. So, the probability that a customer will go to a retail area depends on the probabilities of other options. Huff's model determines attractiveness using size and travel time with the following formula:

$$A_{ij} = \frac{S_j}{T_{ij}^{\lambda}}$$

Where:
Aij = the attractiveness of store j for customers in area i
Sj = the size of store j
Tij = travel time from area i to store j
λ = parameter reflecting propensity to travel. a higher **λ** means that customers are sensitive to travel distance

The usage of travel time as an indicator of attractiveness instead of distance is affirmed in other research as a more realistic measure of a customer's consideration since it includes the encountered barriers during travel (Mayo, Jarvis, and Xander 1988). Furthermore, Huff includes a constant for the type of shopping trips in his model and accounts for competitors, which leads to the probability using the following formula:

$$P_{ij} = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}}$$

where:
Pij =the probability of a consumer at a given point of origin i travelling to a particular retail area j
λ = a parameter which is to be estimated empirically to reflect the effect of travel time on various kinds of shopping trips (Huff, 1964).

Following from the above, the expected number of consumers that are likely to travel to a particular retail area can be determined:

$$E_{ij} = P_{ij} \bullet C_i$$

where:

E_{ij} = the expected number of consumers at i that are likely to travel to retail area j

P_{ij} = the probability of consumers at i that will shop at retail area j

C_i = the number of consumers at i (Huff, 1964)

From this overview of the most important ideas on retail gravitation, some conclusions can be formulated. First, although the models differ, distance and size are included in all models, which shows that these variables are essential when measuring attractiveness. Distance can be formulated in different ways. Reilly assumes an absence of barriers when travelling to a location which causes the theory to fail to capture considerations of obstacles. Replacing distance with travel time, like Huff, makes for a more reliable model since time can be seen as the main barrier. In other research, Huff's model is used to analyze retail turnover. It is found that the model does a fairly good job of predicting turnover and deviates, at a maximum, ten to fifteen percent from the actual values (Egorova et al., 2020). Other scholars found that Reilly's law is better at explaining footfall in rural areas, where inhabitants are more sensitive to the effects of distance. They must travel higher distances since the hometown has a smaller retail offer (Mason & Mayer, 1990; Wagner, 1974).

A limitation of retail gravitation theory is that it fails to include vacation shoppers in its models and cannot predict which part of customers will shop online (Friske & Choi, 2013). Friske & Choi (2013) also argue that 'satisfaction evaluation' should be incorporated in the formula to account for customers' decisions based on subjective opinions. Furthermore, the retail environment and trends changed significantly since the theories of Reilly, Converse and Huff were established, so one could ask if these theories can still explain retail gravitation. The changing environment was already noted in 1994 by Eppli and Benjamin, who also show that many studies include other variables apart from size and distance (Eppli & Benjamin, 1994). From the previous part, the first hypothesis of this research can be formulated:

H1: Size and distance cannot entirely explain retail area performance

2.2 Attractiveness

To understand what makes a retail area attractive, it is important to consider what is meant by a retail area. The definition of a retail area, according to Teller & Reutterer (pp. 127, 2008) is: "sites established consciously, i.e. planned agglomerations such as shopping centers, or unconsciously, i.e. unplanned agglomerations such as shopping streets". The owners' aim is to increase attractiveness, synonymous for gravity, and draw power or preference over other options to consumers, which should lead to sales maximization for tenants. In this research center locations are also considered, which are shops clustered together not in a single building but rather in a shopping street (Teller, 2008).

Mumford at al. (2017) find that size has a significant influence on footfall in a retail area. Large retail areas generate a higher footfall than small retail areas. Additionally, shops cluster together to make use of customers 'trip-chaining', which means that customers visit multiple shops in one trip to a shopping area. Customers do this to reduce their transport costs by having to go to only one place (Koster, Pasidis and van Ommeren, 2019). The phenomenon of trip-chaining can easily be related to retail gravitation theory; when shops cluster together, indicating a bigger size, customers are more likely to go there to reduce transport costs, which can be related to distance. However, according to Dennis, Marsland & Cockett (2002), the expected effect of higher distance is a decrease in attractiveness. So, the size of a shopping area explains the distance a consumer is willing to travel. Koster et al. argue that areas with more shops tend to have a higher footfall. Their research focuses on shopping streets since these are more prominently present in the Dutch retail environment than shopping malls. The authors find that the vacancy rate decreases with 0.35 percentage points when footfall increases by ten percent. So, vacancy rates depend negatively on footfall in shopping streets in the Netherlands.

When looking at other theories that aim at explaining a customer's choice of a shopping area, the law of market areas can also be considered. According to this law, a customer's choice depends on two factors: travel time or distance and price level. So, customers would be willing to travel further when prices are lower. However, the law of market areas is not applicable when centers trade in differing goods (Parr, 1977). Many retail areas trade in differing goods, so the law of market areas is not useful for retail area analysis when the centers are not homogeneous. Nevertheless, the average price level may be a variable explaining attractiveness, especially of discount centers. Hassan & Mishra (2015), on the other hand, find that the success of discount centers can be mainly assigned to convenience of such centers.

Gabswicz & Thisse (1986) show in their research that when shops are homogeneous, the only factor for a customer's decision on where to go is distance. Again, distance is noted as transport costs. More recently, Teller (2008) argues that evolved shopping agglomerations, where ownership is fragmented and the agglomeration evolved unplanned, have declined in attractiveness to customers. On the other hand, created retail agglomerations, which are planned to be agglomerations and are actively managed with less divided ownership, have seen floor space and sales increase. In the paper, the authors try to find what determines the attractiveness of shopping streets and malls. Tenant mix and the presence of an anchor tenant within the agglomeration have a significant influence on the customer's perception of attractiveness as it affects the capability to fulfil the planned task, which is supported in several other studies (Mejia & Benjamin, 2002; Feinberg et al., 2000) Furthermore, the author notes accessibility as an essential factor indicating attractiveness. The results show that tenant mix and atmosphere are important factors explaining the attractiveness of both malls and shopping streets.

Literature on the effect of customer perception on the performance of retail areas is mainly available for shopping malls and individual shops, but not much is available for smaller shopping centers and high streets. Kushwaha et al. (2017) find that customers base their choice of shopping mall on service experience, internal environment, convenience, utilitarian factors, acoustics, proximity and demonstration. A similar research was conducted by Khanna and Seth (2018) for developing cities in India. The authors find that aesthetic ambience, physical infrastructure, hedonic factors, service and convenience, stress relieving, promotions, merchandise, shopping enjoyment and excitement are factors influencing customer choice and can enhance footfall. According to Oppewal and Timmermans (1999), the customer perception of the public space in a shopping center is influenced by the availability of green, maintenance, attractiveness of window displays and the number of street activities.

From this literature review, the second and third hypotheses can be formulated:
H2: *Size and distance have differing effects between types of retail areas*
H3: *Positive customer perception influences performance of retail areas positively*

2.3 Conclusion research question 1

Following the extensive literature review above, it is possible to formulate a conclusion for research question 1:

What are the determinants for the attractiveness of retail areas?

The first part of this literature review discussed the most important retail gravitation theories. According to Reilly, size and distance are the only measures of the attractiveness of a retail area. Later, Huff changed Reilly's model and included travel time instead of distance since it is a more valid measure. However, other existing literature indicates that, while size and distance do play a part in explaining attractiveness, the two variables alone cannot wholly explain the attractiveness of a retail area. On the other hand, some of the variables found in existing literature have a clear relation to size and distance, like shop-clustering or retail agglomerations.

In the second part, the literature on attractiveness is discussed. Koster et al. find a negative relationship between footfall and vacancy rates, suggesting that footfall, a measure of attractiveness, is not only influenced by size and distance but also by vacancy rates. Furthermore, fragmented ownership seems to have a negative impact on the attractiveness of a shopping center since it usually leads to less active management in the shopping center and a lower ability to choose tenants. Finally, The factors influencing customer perception found in the existing literature can mostly be related to atmosphere, tenant mix, accessibility, ambience and service.

In conclusion, the literature review shows that the determinants of the attractiveness of a retail area are; size, distance, or more specifically travel time, vacancy rate, tenant mix, accessibility, price level and atmosphere.

3. DATA & METHODS

3.1 Data

The data necessary to conduct this research is provided within two datasets. The first dataset is from the Koopstromenonderzoek (KSO), a study of retail areas in the Netherlands conducted by I+O Research, BRO and Bureau Stedelijke Planning. The second dataset is provided by Colliers Netherlands. After cleaning and combining both datasets, a total of 491 observations remain.

Dataset Koopstromenonderzoek 2021

The dataset provided by KSO includes data for 618 retail areas in the Netherlands for 2016, 2018 and 2021, with a total number of observations of 1,848. Retail areas are classified according to size and type. The most important variables included in the dataset are turnover, vacancy rates, size and customer valuations.

The dataset includes the location of the retail area. The incorporated area is visualized in figure 3.1. This variable is crucial as location is known as one of the most important explanatory factors of real estate returns. However, the location is included per municipality, leading to a low number of observations for each category. A low number of observations can cause problems with the validity of the results. To create a variable that is more useful to work with, the location variable is recoded according to size. The population for each municipality is added manually to the dataset with data from Eurostat. The included municipalities have over eleven million inhabitants, covering most of the Netherlands (Eurostat, 2022). The variable is not normally distributed, so a new categorical variable is created. Categories are created according to the size of the population. The summary of the newly created variable for *population* category can be found in Appendix I.

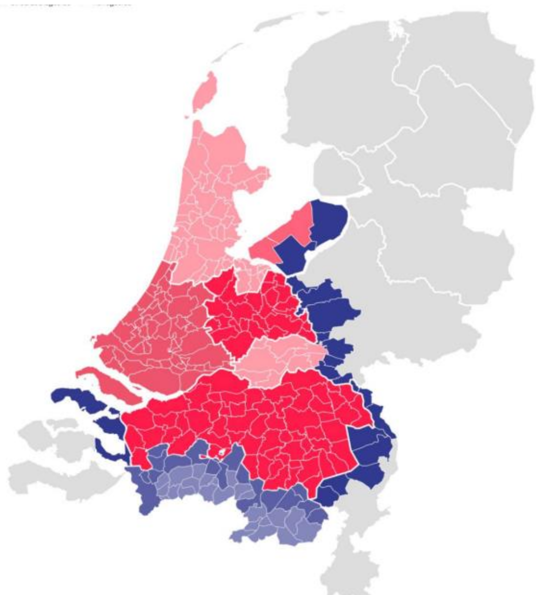

Figure 3.1*:* Focus area included in the dataset (Koopstromenonderzoek, 2021)

The dataset provides observations for turnover per category for groceries, recreational and goal oriented. Turnover for 2021 is generated using the CBS productiestatistiek and Omzetkengetallennotitie 2021, which extrapolates the turnover numbers of 2019 per category based on historical trends (KSO, 2021). Adding all categories together, the total turnover is generated. However, turnover is not available for all retail areas, leading to some observations being dropped. After dropping the missing values for turnover, the dataset contains 1,432 observations. The data is positively skewed, so the variable is transformed to a natural log (*lnturnover)* to create a more normal distribution.

The variable *sqmretail* is measured as the size of the shopping area. The size of the shopping area is the total size of the shopping area measured in square meters. The retail areas are categorized according to size and type as center locations, support locations or furniture malls. Most observations fall in the category *center location between 10,000 and 20,000 square meters* with 185 observations. The least observations are in the category *center locations of over 200,000 square meters*, having only twelve observations. This category is combined with the category for center locations between 100,000 and 200,000 square meters, creating a new category with *center locations bigger than 100,000 square meters,* including 33 observations. The variable is coded from 1 to 16, and descriptive statistics can be found in Appendix I.

The following important variable is vacancy. The vacancy rate in a shopping area is measured for both the stores and for total size with data from Locatus (KSO, 2021). For this research, the variable measuring the vacancy of the total size of the retail area will be used since this is a more realistic representation of actual vacancy. The mean vacancy rate is 6.08 percent, with a standard deviation of 6.29. The minimum value is 0, indicating no vacancy at all, and the highest value is 53 percent. The data are not normally distributed. However, it is impossible to transform to a natural log when there are percentages of 0 in the data (Brooks & Tsolacos, 2010, p. 144). A new variable, *delta_vacancy*, representing the difference between vacancy from period to period is created to solve this issue and the issue of possible endogeneity. Since there are three periods (2016, 2018, 2021), the observations for 2016 are dropped, leading to a total number of observations of 1,034. The new variable has a normal distribution judged by the new histogram.

Furthermore, the variable *avvaluation* is included. The variable includes customer valuations on the offer of cafes and restaurants, green, accessibility, facilities, cleanness, parking, ambiance, safety and tenant mix. Nearly all of these factors were mentioned in the theory about the determinants of the attractiveness of retail areas. The only missing influential factor is the perception of service. The valuation that customers give to a retail area is indicated by a mark between 1 and 10, so it is an ordinal variable. Ordinal variables can be treated as continuous variables in regression analysis. The data are incorporated in the dataset as a number between 0 and 100, where a 7.8 is entered as 78. The variable includes customer valuations on the offer of cafes and restaurants, green, accessibility, facilities, cleanness, parking, ambiance, safety and tenant mix. In addition, valuations based on recreational, goal-oriented and groceries are included, and a new variable for the average of all categories is created. The variable for the average valuation is easier to interpret since it gives the overall valuation of the retail area and accounts for several control variables.

Dataset Colliers

Footfall is the first dependent variable in this research. Footfall data for 2021 included in the dataset are provided by Colliers Spots. The data are available for 491 assets in 2021, so observations for 2018 are dropped from the dataset. The variable is not normally distributed, so a new variable with the logarithm is created; *ln_footfall.* As a robustness check, the test will also be performed with footfall data from Locatus for 2016, 2018 and 2021 (n=77). However, the Colliers Spots data can be trusted to be much more reliable since it is gathered using GPS. Locatus gathers its data by counting passers-by on a random Saturday and then calculates the yearly footfall using a factor. An important note is that in 2021 several lockdowns occurred due to Covid-19. Stores were closed from the January 1st until March 3rd, after which a partial lockdown was in place. Stores were able to fully open again on April 28th. On November 12th, a new partial lockdown started, which became a complete lockdown on December 19th (RIVM, 2022). Therefore, the outcomes will give insight into retail areas' functioning during times of Covid-19 and are less generalizable for non-shock situations.

Avdistance is the second variable explaining attractiveness, according to Reilly. Travel time is the most realistic way to measure distance without relying on heavy assumptions. Reilly's assumption of having no barriers when travelling to a location is unrealistic. When measuring distance as travel time, like in Huff's model, the barriers are included in this time. However, this data is not available, so distance will be measured as the average distance customers travel to the shopping area in kilometers. Also, Huff's model includes the $\boldsymbol{\lambda}$, which is impossible to measure with the available data. Data for distance and footfall are acquired using raw GPS data, which is anonymized. The mean for all observations is 1.8 kilometers. Descriptive statistics for all variables included in the regression, as well as the robustness check, are shown in table 3.1.

Some of the variables found to be influential for determining attractiveness of retail areas were not available to include in this research. Data on travel time, tenant-mix, accessibility, price level and atmosphere are not included. The variables are partly included in the customer valuations but their impact and significance cannot be analyzed in this research.

Table 3.1: Descriptive statistics

Variable	Obs	Mean	Std. dev.	Min	Max
footfall	491	5382118	8759319	120610	124000000
ln_footfall	491	15.05887	0.864568	11.70032	18.63379
turnover	491	51,400,000	65,000,000	3308561	733000000
lnturnover	491	17.4126	0.7508432	15.01202	20.41287
Footfall locatus	77	3992113	2244564	394312	1.13e+07
ln_locfootfall	77	15.03533	.6125495	12.8849	16.24467
sqmretail	491	15506.31	23426.28	600	284900
avdistance	491	1.795438	1.525423	0.32	14.92
avvaluation	491	74.80686	4.12853	56.5	84.5
vacancy_size	491	6.289206	6.693422	0	53
delta_vacancy	491	0.0183299	7.94096	-41	50
SCtype	491	7.773931	4.619371	1	16
population	491	1.712831	1.121687	0	4

The correlation table is shown in table 3.2. There are some relatively high correlated variables, which may cause multicollinearity. The correlation table shows a negative correlation between the distance and valuation, so they move in different directions; when distance increases, the average valuation decreases. As expected, according to Reilly's law, the natural log of footfall is positively correlated with both size and distance, as is the natural log of turnover. Footfall and turnover are highly positively correlated, but this is not a problem since the variables are used in separate regressions. Furthermore, customer valuations have a stronger positive correlation with turnover than footfall. Finally, the delta of vacancy negatively correlates with customer valuations, so customers value a retail area lower when vacancy has increased in the last period. A VIF test will be conducted after the regression and included in Appendix II to test for multicollinearity.

Table 3.2: Correlation matrix main regression

	lnturnover	ln_footfall	sqmretail	avdistance	avvaluation	delta_vacancy
lnturnover	**1.0000**					
ln_footfall	**0.7948**	**1.0000**				
sqmretail	**0.7193**	**0.6056**	**1.0000**			
avdistance	**0.2415**	**0.1086**	**0.4682**	**1.0000**		
avvaluation	**0.2460**	**0.0790**	**0.0688**	**-0.0118**	**1.0000**	
delta_vacancy	**0.0847**	**0.0992**	**0.0893**	**0.0257**	**-0.0665**	**1.0000**

Table 3.3: Correlation matrix robustness test

	ln_locfootfall	lnsize	vacancy_size	lnvaluation
ln_locfootfall	**1.0000**			
lnsize	**0.6271**	**1.0000**		
vacancy_size	**-0.1171**	**-0.2433**	**1.0000**	
lnvaluation	**0.3857**	**0.4915**	**-0.3427**	**1.0000**

The correlation matrix for the robustness test is shown in table 3.3. The correlation table shows some highly and moderately correlated variables, so a VIF test will be included in Appendix II to account for multicollinearity. For example, the natural log of turnover and the size of the retail area are highly positively correlated. The VIF test shows some variables with multicollinearity issues. Multicollinearity causes problems with the interpretation of coefficients and reduces the statistical power of a model (Brooks & Tsolacos, 2010, p.173-174).

3.2 Methods

This thesis aims to find if there is a relationship between the performance of retail areas, footfall and turnover, and attractiveness attributes. Therefore, a method is needed in which multiple independent variables can be inserted. A multiple linear regression is used to analyze the relationship between turnover and footfall and the independent variables. The classical linear regression model is a method to model the relationship between a dependent variable and an independent variable. Multiple linear regression is an extension of the classical linear regression model where the relationship between multiple independent variables and an independent variable can be modelled (Brooks & Tsolacos, pp. 109, 2010). The outcome shows whether there is a relationship between the dependent and independent variables. Furthermore, this method allows predicting the dependent variable based on the coefficients that result from the regression. The regression will be executed with cross-sectional data, which are data at a single point in time at multiple locations. Attractiveness consists of the variables size, distance, vacancy rate and valuation by the customer. These variables are the key independent variables. The control variables are the location effects, consisting of the population category and type of retail area. An ordinary least squares (OLS) approach is used in the regression. A test of the five OLS assumptions that should be met is included in Appendix II. Furthermore, to ensure a transparent research process, the Stata do-file is added in Appendix VII.

Eight models are tested in the first set of regressions, the first four models with the natural logarithm of footfall as the dependent variable and the last four with the natural log of turnover as the dependent variable. Model one tests the retail gravitation theory and includes only size and distance, while model two is a simple linear regression model, which tests the relationship with the delta of vacancy. The third model is also a simple linear regression model where the relationship with the average customer valuation is tested. Finally, the fourth model is the complete model where all attractiveness variables are included, as well as location characteristics in the form of the size of the population and the type of retail area. The same is repeated in models five to eight, with the natural log of turnover as the dependent variable.

The most complete regression equation for both dependent variables can be stated as:

$$Log(FF) = \beta_0 + \beta_1 \times Size + \beta_2 \times Distance + \beta_3 \times \Delta Vacancy$$
$$+ \beta_4 \times valuation\ customer + location\ effects + \mu$$

$$Log(TO) = \beta_0 + \beta_1 \times Size + \beta_2 \times Distance + \beta_3 \times \Delta Vacancy$$
$$+ \beta_4 \times valuation\ customer + location\ effects + \mu$$

Where Log(FF) is the natural log of footfall and Log(TO) is the natural log of turnover. B0 is the constant term, which is the intercept. The error term, μ, is included because it is unrealistic to determine the value of the dependent variable with certainty based on the coefficients. In reality, the data are not entirely generalizable for every asset, and some influential factors are not observable or not included in the data (Brooks & Tsolacos, pp. 75-76, 2010).

It is plausible that the outcomes differ between municipalities and types of retail areas. Therefore, an additional heterogeneity test will be performed where the type of retail area and the size of the population is pooled and interacted. The heterogeneity test will be done for differences in size of the population and differences in the type of retail area. First, the populations are divided into big and smaller populations (<100,000 inhabitants and >100,000 inhabitants) to see if the outcomes are different for the size of the municipality. After that, the difference between size of the retail area is also tested. Data is divided into small, middle, and large retail areas.

Since outcomes can differ between periods, a robustness check will be performed with panel data for the years 2016, 2018 and 2021. However, with 77 observations, the dataset is much smaller than the dataset in the main regression. The data in the robustness test are collected by Locatus and calculated for the entire year, based on one day of observations. The measurement precision is far below the data used in the main regression, which should be considered when interpreting results. For the robustness test, the complete models for turnover and footfall are repeated. However, the data for the average distance traveled by a customer are only available for 2021 and thus not included in the regression. To be able to compare the outcomes the main regression will be repeated without the inclusion of the average distance. Furthermore, to increase the number of observations, the delta of vacancy is transformed back to the actual value so observations for 2016 can be included. To account for the possible problem of endogeneity, robust standard errors are included in the regression.

4. RESULTS

4.1 Regression results

The most important results of the regression are presented in table 4.1 for models 1, 4, 5 and 8. Full results can be found in Appendix III. In the first four models, the dependent variable is the natural logarithm of footfall. In models five to eight the dependent variable is the natural logarithm of turnover. Robust standard errors are used to account for heteroscedasticity. Multicollinearity is tested with a VIF test, showing no problems with multicollinearity and a mean VIF of 2.52 for the footfall model and 3.07 for the turnover model.

Model 1 is the simple model where Reilly's law is tested without including any other independent variables. The outcome shows if Reilly's law is still able to completely explain attractiveness in the modern context. The model includes size of the retail area, and the average distance travelled. Both variables are transformed into a log and show a significant relationship with the natural logarithm of footfall at the one percent level. The coefficient for size of the shopping area is positive, indicating that footfall increases with size. On the other hand, the coefficient for distance is negative, meaning that footfall decreases with the average distance travelled. Reilly's law assumes that size increases the distance a person is willing to travel, so a negative coefficient for distance is not problematic for the outcomes. Furthermore, the R-squared is relatively high at 0.5293, meaning that the model explains 53 percent of variance in the dependent variable. The outcome partly supports Reilly's law, but there are still other explanatory variables that are not included in the model. The regression was also done using a log level method, where the independent variables are not transformed to the natural logarithm. The results can be consulted in Appendix IV and provide the possibility for a different interpretation. For example, a one thousand square meter increase in size leads to a 2.62 percent increase in footfall.

Model 2 includes the delta of vacancy and shows a significant relationship. However, this relationship also differs from the expectation since it is a positive relationship, indicating that when vacancy has increased, compared to the previous period, footfall increases as well. The heterogeneity and robustness tests will further examine the relationship since the outcome counters the evidence found in the literature. Furthermore, the third model, with customer valuations as the dependent variable, is not significant at the five percent level with a p-value of 1.70, and zero in the confidence interval. Therefore, no inferences can be made from this model. Both models are added in Appendix III.

The most complete model for footfall is the fourth, which includes the key independent variables and control variables for population and the type of retail area. Vacancy, customer valuations and some of the categories are not significant at the five percent level, which is not unusual in a regression with many categories. The positive relationship with size has a lower coefficient, and the negative relationship with distance is also more minor, which means that the effects of size and distance are smaller when other variables are included. For example, an increase of one percent in size would increase footfall by 0.59 percent. This percentage seems small, but when measuring for an increase of 10 percent, footfall would increase by 5.9 percent. The R-squared is high at 0.7044, so the model explains 70 percent of variance in the natural logarithm of footfall. Although the model explains a high percentage, some explanatory variables are still missing. Looking at the log-level model, an increase of 1,000 square meters in size leads to an increase of 2.49 percent in turnover.

In model 5, Reilly’s law is tested for turnover. Again, a significant relationship is found for both variables. An increase in size of one percent leads to a 0.66 percent increase in turnover. The coefficients are smaller than in the model with footfall as the dependent variable, so the relationship is stronger for footfall. However, the R-squared is bigger at 0.6782, which is higher than in the footfall model. So, Reilly’s law can better explain turnover than footfall.

Model 6, where the natural log of turnover is regressed against the delta of vacancy, is not significant, so that no inferences can be made based on this model. The model where customer valuation is included is significant at the one percent level. If the average valuation of the retail area increases by one percent, the turnover per square meter increases by 3.28 percent. Retail areas that are valued highly by customers have higher turnovers. However, only 0.7 percent of variance in the natural log of turnover per square meter is explained by this model.

Finally, model 8 is the complete model with turnover as the dependent variable. Size and distance have similar outcomes to the footfall model. Again, the delta of vacancy is not significant and will be studied in further tests. The categories for both population and the type of retail area show high coefficients. For example, when the population falls in the biggest category (over 250,000 inhabitants), turnover per square meter is 43 percent higher compared to the smallest category. Coefficients grow with the size of the population, which was expected. The base category for the type of retail area is a center location of 10,000 to 20,000 square meters, which is the mode in the dataset. Most categories do not show a significant relationship with the natural log of turnover, so that no inferences can be made. However, 83 percent of variance in turnover is explained by the model, which is a high percentage.

4.2 Heterogeneity

The results may differ between areas and types of retail areas. To find if there are significant differences between categories, the regression is repeated with an interaction between the type of retail area and the size of the population. The results are shown in table 4.2. The outcomes are similar to the main regression. Again, a significant relationship is found between size and distance. The coefficients of size are smaller than in the main regression, and the coefficients of distance are similar. The relationship

Table 4.1: Regression results main regression

	(1) ln_footfall	(4) ln_footfall	(5) lnturnover	(8) lnturnover
lnsize	**0.708*** (19.92)**	**0.588*** (6.32)**	**0.661*** (27.10)**	**0.717*** (7.85)**
lndistance	**-0.395*** (-5.50)**	**-0.208** (-3.26)**	**-0.209*** (-5.01)**	**-0.0938* (-2.44)**
lnvaluation		**-0.594 (-1.40)**		**1.993*** (6.78)**
delta_vacancy		**-0.00321 (-1.15)**		**0.000339 (0.15)**
Center location 10,000-20,000 sqm		**0 (.)**		**0 (.)**
Center location >100,000 sqm		**0.904*** (4.01)**		**0.303 (1.50)**
Center location 20,000-40,000 sqm		**0.0709 (0.78)**		**0.0532 (0.70)**
Center location 40,000-60,000 sqm		**0.127 (0.91)**		**0.0389 (0.30)**
Center location 5,000-10,000 sqm		**-0.131 (-1.27)**		**0.0124 (0.15)**
Center location 60,000-100,000 sqm		**0.402* (2.20)**		**0.0754 (0.46)**
Center location <5,000 sqm		**-0.0673 (-0.41)**		**0.0602 (0.51)**
Supportive 10,000-20,000 sqm		**0.113 (0.87)**		**0.107 (1.28)**
Supportive 2,500-5,000 sqm		**-0.205 (-1.36)**		**0.303* (2.29)**
Supportive 5,000-10,000 sqm		**-0.00223 (-0.02)**		**0.244** (2.82)**
Supportive <2,500 sqm		**-0.264 (-1.38)**		**0.407* (2.29)**
Supportive >100,000 sqm		**0.0705 (0.44)**		**0.0107 (0.09)**
Furniture mall 20,000-40,000 sqm		**-0.912*** (-5.94)**		**-0.497*** (-3.85)**
Furniture mall <20,000 sqm		**-1.085*** (-6.94)**		**-0.625*** (-6.19)**
Furniture mall > 40,000 sqm		**-0.581* (-2.52)**		**-0.276 (-1.42)**
Population <20,000		**0 (.)**		**0 (.)**
Population 20,000-50,000		**0.0962 (1.07)**		**0.118 (1.89)**
Population 50,000-100,000		**0.238* (2.28)**		**0.172* (2.41)**
Population 100,000-250,000		**0.295** (2.77)**		**0.304*** (4.09)**
Population 250,000+		**0.451** (2.78)**		**0.437*** (4.72)**
_cons	**8.779*** (28.23)**	**12.31*** (6.24)**	**11.49*** (53.42)**	**13.13** (2.38)**
R-squared	0.5293	0.7044	0.6782	0.8263
N	**491**	**491**	**491**	**491**

t statistics in parentheses
robust standard errors

* p<0.05, ** p<0.01, *** p<0.001

between the average customer valuation and footfall is insignificant, while the relationship with turnover shows a significant outcome at the 0.1 percent level. Finally, vacancy is not significant for both dependent variables. The most interesting part is the outcome of the interaction variables. The base category is a small retail area in a population of under 100,000 inhabitants. Nearly all categories are significant, except for a small retail area in a large population and footfall.

Starting with footfall, a medium-sized retail area in a small population experiences a 38.5 percent higher footfall compared to a small retail area in a small population. A large retail area in a small population scores even higher and has a 56.7 percent higher footfall compared to the base category. Looking at large populations of over 100,000 inhabitants, the result for small retail areas is insignificant. However, the other categories show significant results with relatively high coefficients. A medium-sized retail area in a large population experiences a 60.4 percent higher footfall, and large retail areas a 121.5 percent higher footfall. The difference between the coefficient in a large retail area in a small population and a large retail area in a large population is considerable.

Table 4.2: Regression results heterogeneity test

	(9) ln_footfall	(10) lnturnover
lnsize	**0.471*****	**0.497*****
	(8.84)	**(13.66)**
lndistance	**-0.383*****	**-0.205*****
	(-5.77)	**(-5.54)**
lnvaluation	**-0.243**	**2.149*****
	(-0.57)	**(7.11)**
delta_vacancy	**0.00390**	**0.00196**
	(1.36)	**(0.83)**
Small retail area # population <100,000	**0**	**0**
	(.)	**(.)**
Small retail area # population>100,000	**0.0437**	**0.217*****
	(0.48)	**(3.98)**
Medium retail area # population<100,000	**0.385*****	**0.228*****
	(4.29)	**(3.63)**
Medium retail area # population>100,000	**0.604*****	**0.461*****
	(6.30)	**(7.86)**
Large retail area # population<100,000	**0.567****	**0.381****
	(2.70)	**(2.74)**
Large retail area # population >100,000	**1.215*****	**0.939*****
	(7.15)	**(7.77)**
_cons	**11.65*****	**3.468****
	(5.98)	**(2.62)**
R-squared	0.5953	0.7669
N	**491**	**491**

t statistics in parentheses
robust standard errors
* p<0.05, ** p<0.01, *** p<0.001

For turnover, all categories show a significant result. An interesting outcome is that a small retail area in a large population has a 21.7 percent higher turnover than the base category. This result shows that turnover is sensitive to the size of the municipality that the retail area is in. The same pattern is seen in the result for medium and large retail areas. All retail areas in larger populations have a higher coefficient than equivalent retail areas in smaller populations.

4.3 Robustness

In the robustness checks panel data will be used, which are data at multiple points in time at multiple locations. The difference between the data is because the footfall data from Colliers Spots are only available for 2021 but for 491 assets, while the Locatus data are available for multiple years but fewer assets. The Locatus data includes 22 observations for both 2016 and 2018 and 33 observations for 2021. The robustness check aims to find if the regression results are the same when including multiple periods. However, the average distance is only available for 2021, so this variable will not be included. Because of this, it is impossible to test Reilly's law in the robustness checks. To compare the different models, the main regression is repeated without the inclusion of average distance in models 13 and 14. Descriptive statistics for models 11 and 12 can be found in Appendix V. The most important results of the robustness test are shown in table 4.3, and full results are added in Appendix VI. Models 11 and 12 were executed using Locatus data, and models 13 and 14 with Colliers Spots data. Vacancy is transformed back to the actual value because it includes the year 2016, which creates the possibility to compare a more extended period. This does, however, mean that there is an increased chance of endogeneity, so robust standard errors are included.

Table 4.3: Regression results robustness test

	(11) Ln_locfootfall	**(12) Ln_locfootfall**	**(13) Ln_footfall**	**(14) Ln_footfall**
lnsize	**0.574*** **	**-0.132**	**0.580*** **	**0.546*** **
	(9.71)	**(-0.45)**	**(18.27)**	**(6.08)**
lnvaluation		**0.413**		**-0.540**
		(0.28)		**(-1.24)**
vacancy_size		**-0.00123**		**0.00121**
		(-0.10)		**(0.28)**
Location effects	NO	YES	NO	YES
_cons	**8.723*** **	**14.77**	**9.796*** **	**12.36*** **
	(13.23)	**(1.83)**	**(34.99)**	**(6.17)**
R-squared	0.3933	0.5332	0.4737	0.6915
N	**77**	**77**	**491**	**491**

t statistics in parentheses
robust standard errors
* p<0.05, ** p<0.01, *** p<0.001

A positive significant relationship is found between size and the natural logarithm of footfall. The result is very similar to that from the Colliers Spots data. A one percent increase in size leads to a 0.57 percent increase in footfall and 0.58 percent in model 13. Model 13 is slightly better at explaining variance in the dependent variable with an R-squared of 0.47, compared to 0.39 in model 11. Looking at the complete models 12 and 14, there is a considerable difference between the outcomes. In model 12, none of the independent variables has a significant result, and neither does the constant. The model did not find a significant relationship between footfall and any of the independent variables, which is an evident difference from the main model. Since no significance was found, no inferences can be made based on this model.

5. DISCUSSION

In this chapter, the results will be discussed in-depth, and related to the hypotheses. The regression results show that there is indeed a relationship between the attractiveness of retail areas and their performance. The leading theory tested in this research is Reilly's law of retail gravitation and the adjusted theory by Huff, where travel time is included instead of crow fly distance. In the best scenario, the theory by Huff would be tested since it is not plausible that barriers do not influence customers during their travel. However, the data necessary to test this theory are not available, so it is impossible to test Huff's theory in the way it was meant to be tested. The crow fly distance data are available, so Reilly's original law was tested in this research. Reilly finds that size and distance are the primary drivers of retail gravitation. In the first regression model, a significant positive relationship is found between size and footfall, while a negative significant relationship is found between the dependent variables and average distance. So, the size of a retail area increases footfall, while the average distance a customer must travel to the retail area decreases footfall. This outcome is in line with the existing literature by Dennis, Marsland & Cockett (2002), where it was found that an increase in distance leads to a decrease in attractiveness. The model explains 53 percent of variance in footfall, meaning that some explanatory variables are missing, and we can say, based on the sample, that Reilly's law of retail gravitation is not entirely true. Although a significant part of variance in footfall is explained by size and distance, other factors also influence retail gravitation. The result aligns with the expectations based on the first hypothesis and the existing literature by Friske & Choi (2013) and Eppli & Benjamin (1994), who found that not all explanatory variables are included in Reilly's law, and that the explanatory factors change due to the dynamic retail environment.

Overall, retail gravitation theory forms a firm basis in explaining both footfall and turnover, but other factors also influence the customer's decision. For example, the theory does not make a distinction between countries and thus implies that each countries' retail market functions similarly. Since the geographical layout of the United States is incomparable to that of the Netherlands, it is unlikely that this does not influence a customers' decision. A possible solution can be to include a term in the equation that accounts for geographical differences. Interestingly, the turnover model has a higher R-squared than the footfall model and can explain 68 percent of variance in turnover. So, Reilly's law is better at explaining turnover than footfall, which is an outcome that would be in line with the existing theory of tests of the Huff model. In other research by Egorova et al., (2022), Huff's model was able to precisely predict turnover in a retail area. The coefficients are like the footfall model, so turnover reacts similarly to the impact of size and distance.

Furthermore, the heterogeneity test in table 4.2 shows that retail areas in bigger municipalities tend to have higher footfall. The coefficients for the interaction with retail areas in municipalities with over

100,000 inhabitants are above those for retail areas in municipalities with under 100,000 inhabitants. The effect is a bit weaker for turnover, but for similar retail areas, the ones in bigger municipalities have higher coefficients. This means that retail gravitation theory is not uniform for all types of retail areas, and hypothesis two can be accepted. The outcome agrees with the literature by Gabswicz & Thisse (1986). They find that distance is the only factor influencing a customer's decision of where to go, only if retail areas are homogeneous, which is not the case.

The relationship between the delta of vacancy and the dependent variables is not evident. In model 2, a significant relationship is found, but the coefficient is positive, which means that footfall would increase when vacancy increased compared to the previous period. This outcome is contrary to the expectation based on the existing literature, where a higher vacancy decreases attractiveness and, consequently, the performance of the retail center. Overall, no clear relationship between the difference in vacancy and performance of retail areas can be found in this research, as only one model has a significant outcome. The outcome is not in line with the literature by Koster, Pasidis and van Ommeren (2019), who found that a higher footfall leads to lower vacancy rates. Possible explanations for this opposite outcome are that there might be an error in the data, or the relationship simply does not exist in the selected sample. When looking at the correlation table, a positive relationship between the delta of vacancy and the performance variables is found. Because of this, it is probable that the sample is not able to capture this relationship due to the size and timeframe of the dataset. Furthermore, the time frame may be too short to find the effects of a higher vacancy on performance. The effects of increasing vacancy may not become visible in the year it occurs but at a later time. To further test this relationship, a robustness test was performed. However, the robustness check, where data for a longer time frame is used, shows no significant relationship between vacancy and performance either. The correlation table for the robustness test shows the expected relationship between footfall and vacancy, which is negative. In this case, the absence of a relationship may also be due to the small number of observations and the quality of the data.

No significant relationship was found between the average valuation of a customer and footfall in the main regression and the other tests. Because of this, hypothesis three cannot be accepted in terms of footfall. The relationship is evident in the models with turnover as the dependent variable. Model 7 shows a 32.8 percent increase in turnover when the valuation of a customer increases by ten percent. However, it should be noted that the customer valuations explain only 0.7 percent of variance in turnover in model 7 so the model does not explain much of the variance in turnover. When other variables are added to the model, the coefficient decreases. In model 8, which is the most complete model, a ten percent increase in the valuation of customers increases turnover by 19.9 percent. Overall, the hypothesis that a positive customer perception positively influences performance can be accepted for turnover. This result follows the expectation based on the literature by Friske & Choi (2013), who argue that satisfaction should be incorporated in retail gravitation formulae. The outcome shows the importance of customer perception when managing a retail area. The goal is to generate a turnover to be able to exploit the retail area, so paying attention to ambiance, accessibility and other factors that influence a customer's perception is essential.

As found in the literature review, adding other variables in the model significantly increases the R-squared. The most complete model explains seventy percent of variance in footfall and 83 percent of variance in turnover. In this model, the key independent variables and location-related variables are included. The population of the municipality the retail area is situated in greatly influences both footfall and turnover. The heterogeneity test shows that bigger retail areas have higher footfall and turnovers than smaller retail areas. This outcome is in line with the theory by Mumford et al. (2017), who find that larger retail areas tend to have a higher footfall than smaller retail areas. A small retail area in a small population generates the least footfall and turnover compared to the other categories. This outcome makes sense since a larger retail area can use more of its space as retail space. Furthermore, in municipalities with more inhabitants, the footfall is higher. A logical explanation can be found in the variable for the average distance a customer has to travel. The coefficient is negative, so when distance increases, footfall and turnover decrease. In large municipalities, there are more people within a shorter distance, so the catchment is bigger than in smaller municipalities.

The robustness test has an interesting outcome when looking at the increase in R-squared when adding location effects. In the most complete footfall model, the R-squared is 53 percent, which is lower than in the most complete models in the main regression and the heterogeneity test. The cause of this difference may be because of the time effects included in this model. Also, the data are for a much smaller dataset and are prone to measurement error. Scholars in the Netherlands widely use the footfall data used in the robustness test. However, the data have significant differences in values compared to the data provided by Colliers Spots. For example, a total of 5,329,461 people visited the city center of Amsterdam in 2021, according to Locatus, and 123,751,800 according to Colliers Spots. This immense difference may indicate that previous research using Locatus data may not be representative as the data are far from accurate. With the currently available technologies, it is impossible to count footfall with a hundred percent accuracy. Therefore, the outcomes from the robustness test will not be used to make any suggestions for further research or to assess the impact of the outcomes on the retail real estate market.

The research encountered some limitations due to data availability. The results would have been more useful if distance had been measured as travel time. Crow fly distance is not realistic, and barriers encountered during travel influence willingness to travel, according to Huff. However, it was not possible to get access to such data with the available resources. Furthermore, the data by Colliers Spots are very reliable as it is measured throughout the year. If the data had been available for multiple years, a comparison could have been made between pre-Covid and post-Covid. The robustness test aimed to solve this problem and find if the outcomes were similar when the analysis included multiple years. In the robustness test, data by Locatus was used for a smaller number of retail areas for multiple years. As mentioned before, the quality of the data is questionable. When comparing observations, considerable differences are found between the Locatus data and the data from Colliers Spots. Because of this, the interpretation of the outcomes based on Locatus data should be done with consideration. Finally, in this research, turnover and footfall were used to measure retail gravitation, while other variables may also be used as a measure of performance. Thus, this research tested whether retail gravitation theory, or attractiveness, can explain footfall and turnover in a retail area. Additionally, some of the determinants of attractiveness that were found in the existing literature were not available for this research. Adding these variables may increase the R-squared and influence the coefficients.

6. CONCLUSION

This research aimed to gain insight into the relationship between attractiveness and performance of a retail area. The main research question in this study is "*To what extent does the attractiveness of retail real estate in the Netherlands influence the performance of retail areas?*".

Attractiveness is essential in explaining the performance of retail areas and explains a great part of variance in both footfall and turnover. Reilly's law of retail gravitation was tested in this study and was found not to be completely accurate. Although size and distance establish a relatively high R-squared, there are still variables missing in Reilly's law. With this finding, the first hypotheses in this research, size and distance cannot entirely explain retail area performance, can be accepted. For example, location factors add significant explanatory power to the model. Furthermore, the results differ between size of the population and the type of retail area, in line with hypothesis two. Therefore, the theory should account for such differences. Retail areas of all types in bigger municipalities generate a considerably larger turnover compared to retail areas in the smallest municipalities.

Additionally, one of the most important and surprising outcomes is the influence of customer valuations on turnover. A one percent higher customer valuation leads to a 1.99 percent increase. With this outcome, the third and final hypothesis can be accepted: positive customer perception influences performance positively. The importance of customer valuations shows that size and distance alone cannot entirely explain attractiveness. In future theories, customer satisfaction should be considered when determining turnover and footfall.

To conclude, to what extent does attractiveness influence the performance of retail areas? In this research, Reilly's law of retail gravitation was tested, so his consideration of attractiveness is used while answering the research question. If attractiveness is seen as only size and distance, as Reilly does, it can explain performance to some extent, but other variables are missing when trying to predict footfall and turnover. To be exact, attractiveness can explain 53 percent of variance in footfall and 68 percent of variance in turnover. The difference may be in the type of data used in this research. For instance, the difference may be found within travel time instead of average distance. Other effects may be the economic situation within the area, tourism, or even the weather.

Retail gravitation theorists claim to be able to predict both footfall and turnover from attractiveness. So, attractiveness influences performance to a considerable extent, but other variables also have an influence. So, the error when trying to explain performance may be in Reilly's definition of attractiveness itself. The explanatory power of the model increases considerably when adding location-related variables. The location obviously contains some gravitational power, so when location effects are added to the definition of attractiveness, the influence of attractiveness on performance increases to seventy percent for footfall and 82 percent for turnover. Thus, maybe Reilly's law can continue to exist with alterations to fit the modern mechanism of retail real estate. Some of the additional explanatory variables were already discovered in this research. So, a future theory should at least include customer satisfaction and a way to account for the size of the municipality and the type of retail area.

The results add to the retail market knowledge and can be integrated into practice. The considerable effect of customer valuations indicates that retailers and asset managers should focus on creating the best experience for customers based on the elements included in the variable. For example, a recent news article states that the inner city of Utrecht is doing better than the inner cities of Amsterdam, Rotterdam and the Hague (PropertyNL, 2022). Looking at the dataset used in this research, Utrecht scored 82, which is above the mean of 74.8. Amsterdam scores 76.6, and Rotterdam and the Hague score 77. Comparing this to the shopping center in de Bogaard, a retail area that has not been performing well (PropertyNL, 2019), we see that the average valuation is below the mean, at 65. So, based on these two articles, it is also visible from practice that retail areas with higher customer valuations perform better.

The results are helpful to understand the functioning of retail areas. The main goal of retailers is to generate turnover. With an R-squared of 83 percent, the variables explain a great part of the variation in turnover. Retailers, developers and asset managers can use this information when establishing or adjusting retail areas. This research focused on 2021, a year with great shocks in the retail sector. Further research could focus on a larger time frame to better understand retail mechanisms. In this way, it is also possible to find if consumers tend to shop closer to home since Covid-19, like KSO 2021 argues, and how significant the difference in average distance is compared to previous years. Furthermore, the height of the coefficient for customer valuations is an interesting outcome that could be researched further. In this research, the variable was used as an average of several categories, so the results do not show which categories are more influential. When including the different categories it can be found which factors within customer perception are more and less influential on performance of retail areas. Finally, where this research looked at all types of retail areas, it would also be interesting to focus on one type of retail area as the heterogeneity test found significant differences between the sizes of retail areas and the municipalities in which the retail areas are located.

REFERENCES

BouwInvest, (2020). *Real Estate Market Outlook 2021-2023*. Accessed on June 28th, 2022 from: https://www.bouwinvest.com/news/latest-news/2020/covid-19-demand-shock-won-t-derail-favourable-long-term-prospects-for-dutch-real-estate-bouwinvest-2021-2023-outlook/

BouwInvest. (2021). *Real Estate Market Outlook 2022-2024.* Accessed on June 28th, 2022 from: https://www.bouwinvest.nl/nieuws/publicaties/2021/market-outlook-2022-2024/

Brooks Ch., Tsalacos S., (2010). *Real Estate Modelling and Forecasting*, Cambridge University Press.

CBRE (2022). *Retail outlets in transition, logistics real estate breaks records.* Accessed on June 28th, 2022 from: https://news.cbre.nl/cbre-retail-outlets-in-transition-logistics-real-estate-breaks-records/

CBS (2022). *Consumptie huishoudens groeit met ruim 11 procent in maart.* Accessed on May 25th, from: https://www.cbs.nl/nl-nl/nieuws/2022/20/consumptie-huishoudens-groeit-met-ruim-11-procent-in-maart

Converse, P.D. (1949). New laws of retail gravitation. *Journal of Marketing*, 14 (January), 379- 384.

Cushman & Wakefield (2022). *The Netherlands marketbeat.* Accessed on June 28th, 2022 from: https://www.cushmanwakefield.com/nl-nl/netherlands/insights/netherlands-marketbeat

Dennis, C., Marsland, D., & Cockett, T. (2002). Central place practice: shopping centre attractiveness measures, hinterland boundaries and the UK retail hierarchy. *Journal of Retailing and Consumer Services*, *9*(4), 185-199.

Egorova, A. O., Islamova, G. I., Kuryleva, O. I., Sineva, N. L., & Yashkova, E. V. (2020). Evaluation of Trade Turnover of the Shopping Center with the Help of the Huff Model—Theory and Practical Aspects. In *Growth Poles of the Global Economy: Emergence, Changes and Future Perspectives* (pp. 767-774). Springer, Cham.

Enoch, M., Monsuur, F., Palaiologou, G., Quddus, M. A., Ellis-Chadwick, F., Morton, C., & Rayner, R. (2021). When COVID-19 came to town: Measuring the impact of the coronavirus pandemic on footfall on six high streets in England. *Environment and Planning B: Urban Analytics and City Science*, 23998083211048497.

Eppli, M. J., & Benjamin, J. D. (1994). The Evolution of Shopping Center Research: A Review and Analysis. *The Journal of Real Estate Research*, *9*(1), 5–32. http://www.jstor.org/stable/44095533

Eurostat (2022). *Population database*. Accessed on April 13th, 2022 from https://ec.europa.eu/eurostat/web/population-demography/demography-population-stock-balance/database

Feinberg, R., Stanton, J., Keen, C., Kim, I.-S., Hokama, L., de Ruyter, K., (2000). Attraction as a determinant of mall choice. *Proceedings of the 7th International Conference on Retailing and Services Science,* European Institute of Retailing and Services Science, Eindhoven

Friske, W. and Choi, S., (2013). Another look at retail gravitation theory: History, analysis, and future considerations. *Academy of Business Disciplines Journal*, *5*(1), pp.88-106.

Gabszewicz, J. J. & Thisse, J.F. (1986). On the nature of competition with differentiated products. *Economic Journal* 96, 160-172

Hasan, A., & Mishra, S. (2015). Key Drivers Influencing Shopping Behavior in Retail Store. *IUP Journal of Marketing Management*, *14*(3).

Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing*, 28, 34-38.

Ismail, S. (2006). Spatial autocorrelation and real estate studies: A literature review. *Malaysian Journal of Real Estate, 1*(1), 1-13.

Khanna, P., & Seth, S. (2018). Consumer perception towards shopping malls: Evidence from a Tier II city. *Indian Journal of Marketing*, *48*(4), 47-59.

Koopstromenonderzoek 2021 (2021). *Onderzoeksverantwoording.* Accessed on April 13th, 2022 from: https://kso2021.nl/resultaten/

Koster, H. R., Pasidis, I., & van Ommeren, J. (2019). Shopping externalities and retail concentration: Evidence from Dutch shopping streets. *Journal of Urban Economics*, *114*, 103194.

Koster, H., Tang, C. K., van Ommeren, J., & Bras, N. (2021). Covid-19 and Shopping Streets. *Centre for Economic Policy Research.* https://cepr-org.vu-nl.idm.oclc.org/active/publications/discussion_papers/dp.php?dpno=15995

Kushwaha, T., Ubeja, S., & Chatterjee, A. S. (2017). Factors influencing selection of shopping malls: an exploratory study of consumer perception. *Vision*, *21*(3), 274-283.

Locatus. (2020). *Na jaren van herstel neemt de leegstand in Nederland weer flink toe.* Accessed on November 15th, 2021 from: https://locatus.com/blog/na-jaren-van-herstel-neemt-de-leegstand-weer-flink-toe/

Locatus (2022). *Spectaculaire daling winkelleegstand door afname retailpanden. Accessed* on February 13th, 2022 from: https://locatus.com/blog/spectaculaire-daling-winkelleegstand-door-afname-retailpanden/

Mason, B. J. and Mayer W. L. (1990). Modern Retailing: *Theory and Practice,* Homewood, IL, BPI-Irwin, 679-681.

MDBS. (2014). *Winkelstraten leegstand.* Accessed on November 15th, 2021 from: https://www.mdbs.nl/projecten/winkelstraten-leegstand#:~:text=De%20oorzaken%20van%20die%20groeiende,winkels%20voor%20te%20weinig%20mensen

Mejia, L. C., & Benjamin, J. D. (2002). What Do We Know About the Determinants of Shopping Center Sales? Spatial vs. Non-Spatial Factors. *Journal of Real Estate Literature, 10*(1), 3–26. http://www.jstor.org/stable/44103432

Mejia, L. C., & Eppli, M. J. (2003). Inter-center retail externalities. *The Journal of Real Estate Finance and Economics*, *27*(3), 321-333.

Mumford, C. L., Parker, C. R., Ntounis, N., & Dargan, E. (2017). A clustering study to verify four distinct monthly footfall signatures: A classification for UK retail centres. *Institute of Place Management*.

Nanda, A., Xu, Y. and Zhang, F., (2021). How would the COVID-19 pandemic reshape retail real estate and high streets through acceleration of E-commerce and digitalization?. *Journal of Urban Management,* 10(2), pp.110-124

NOS. (2020). *Leegstand winkelpanden neemt komende jaren hard toe.* Accessed on November 15th, 2021 from: https://nos.nl/artikel/2359237-leegstand-winkelpanden-neemt-komende-jaren-hard-toe

NOS (2022a). *Meer winkels: vooral toename bakkers, supermarkten en slijterijen.* Accessed on June 28th, from: https://nos.nl/artikel/2427328-meer-winkels-vooral-toename-bakkers-supermarkten-en-slijterijen

NOS (2022b). *Volgens deze winkeliers is van een depressie nog lang geen sprake.* Accessed on May 25th, from: https://nos.nl/nieuwsuur/artikel/2426020-volgens-deze-winkeliers-is-van-een-depressie-nog-lang-geen-sprake

Oppewal, H., & Timmermans, H. (1999). Modeling consumer perception of public space in shopping centers. *Environment and behavior*, *31*(1), 45-65.

Parr, J. B. (1997). The law of retail gravitation: insights from another law. *Environment and Planning A*, *29*(8), 1477-1492.

Philp, S., Dolega, L., Singleton, A., & Green, M. (2021). Archetypes of Footfall Context: Quantifying Temporal Variations in Retail Footfall in relation to Micro-Location Characteristics. *Applied Spatial Analysis and Policy*, 1-27.

PropertyNL. (2019). *Ambitieus plan voor In de Bogaard.* Accessed on June 24th, 2022 from: https://propertynl.com/Nieuws/Ambitieus-plan-voor-In-de-Bogaard-2-0/b1e567b9-4fc7-4232-a632-d0eb31ccf3ad

PropertyNL. (2022). *Utrecht centrum behoudt positie.* Accessed on June 24th, 2022 from: https://propertynl.com/Nieuws/Utrecht-centrum-behoudt-positie/5fc2c68e-5123-491f-8d31-b3d75e895d70

Reilly W. J., (1931). *The Law of Retail* Gravitation (New York; The Knickerbocker Press, 1931).

RIVM (2022). *Tijdlijn van coronamaatregelen.* Accessed on May 30th, 2022 from: https://www.rivm.nl/gedragsonderzoek/tijdlijn-maatregelen-covid

Talen, E., & Park, J. (2021). Understanding Urban Retail Vacancy. *Urban Affairs Review*, 10780874211025451.

Teller, C., (2008). Shopping streets versus shopping malls–determinants of agglomeration format attractiveness from the consumers' point of view. *The International Review of Retail, Distribution and Consumer Research*, *18*(4), pp.381-403.

Teller, C., & Reutterer, T. (2008). The evolving concept of retail attractiveness: what makes retail agglomerations attractive when customers shop at them?. *Journal of Retailing and Consumer Services*, *15*(3), 127-143.

Vastgoedmarkt. (2022). *Hoe waardedaling grote winkelcentra tot stilstand komt.* Accessed on February 17th, from: https://www.vastgoedmarkt.nl/retail/nieuws/2022/02/hoe-waardedaling-grote-winkelcentra-tot-stilstand-komt-101172695

Wagner, W. B. (1974). An Empirical Test of Reilly's Law of Retail Gravitation, *Growth and Change*, 5 (3), 30-35.

Zhang, S., (2018). How Is Retail Rents Affected by Retail Mix? Evidence from the Netherlands (No. eres2018_319). *European Real Estate Society (ERES).*

Zhang, J. Z., Chang, C.W., & Neslin, S. A. (2022). How Physical Stores Enhance Customer Value: The Importance of Product Inspection Depth. *Journal of Marketing*, *86*(2), 166–185.

Zhang, S., Van Duijn, M., Van der Vlist, A.J. (2019). The external effects of inner-city shopping centers: Evidence from the Netherlands. *Journal of Regional Science*, 60(4), 538-611.

APPENDIX I: Descriptive statistics categorical variables

Table I.1: Descriptive statistics population

Population categories	Freq.	Percent	Cum.
up to 20.000	**58**	**11.81**	**11.81**
20.000 to 50.000	**194**	**39.51**	**51.32**
50.000 to 100.000	**101**	**20.57**	**71.89**
100.000 to 250.000	**107**	**21.79**	**93.69**
250.000+	**31**	**6.31**	**100.00**
Total	**491**	**100**	

Table I.2: Descriptive statistics type of retail area

Benchmarkgroep	Freq.	Percent	Cum.
Center location 10.000-20.000 m2	**72**	**14.66**	**14.66**
Center location from 100.000 m2	**12**	**2.44**	**17.11**
Center location 20.000-40.000 m2	**44**	**8.96**	**26.07**
Center location 40.000-60.000 m2	**15**	**3.05**	**29.12**
Center location 5.000-10.000 m2	**52**	**10.59**	**39.71**
Center location 60.000-100.000 m2	**9**	**1.83**	**41.55**
Center location up to 5.000 m2	**26**	**5.30**	**46.84**
Supportive 10.000-20.000 m2	**39**	**7.94**	**54.79**
Supportive 2.500 – 5.000 m2	**67**	**13.65**	**68.43**
Supportive 5.000 – 10.000 m2	**48**	**9.78**	**78.21**
Supportive <2.500 m2	**31**	**6.31**	**84.52**
Supportive from 20.000 m2	**11**	**2.24**	**86.76**
Furniture mall 20.000-40.000 m2	**23**	**4.68**	**91.45**
Furniture mall < 20.000 m2	**25**	**5.09**	**96.54**
Furniture mall > 40.000 m2	**17**	**3.46**	**100.00**
Total	**491**		**100**

APPENDIX II: OLS assumptions

OLS assumptions

The classical linear regression model uses ordinary least squares as the estimation technique. To use ordinary least squares, some assumptions should be met.

1. Average value of the errors is zero $E(ut) = 0$
2. Constant error variance $var(ut) = \sigma 2 < \infty$
3. Covariance between error terms is zero $cov(ui, uj) = 0 \text{ for } i \neq j$
4. Regressors are not correlated with the error term $(cov(ut, xt) = 0)$
5. Errors are normally distributed (Brooks & Tsolacos, 2010, pp. 136)

Assumption one is about the average of the error terms being zero. The first assumption is always met when a constant term is included in the regression, which is the case in this research. Including a constant term ensures that the regression line is not forced through the origin (Brooks & Tsolacos, 2010, p. 137-138).

Furthermore, the second assumption is about homoscedasticity. When the error variance is not constant the problem of heteroscedasticity occurs. The Breusch-Pagan test shows that there is a problem with heteroscedasticity in the data for the model with footfall (p=0.0000) as well as for the model with turnover (p=0.0489). However, when heteroscedasticity is present, the model still gives unbiased coefficients. The standard errors may be wrong, but it is still possible to make inferences based on the coefficients. To further analyze the heteroscedasticity problem a residual versus fitted plot is generated, shown in figure II.1. The plot on the left shows the residuals for the footfall model and on the right for the turnover model. Robust standard errors are used in the regression to account for heteroscedasticity (Brooks & Tsolacos, pp. 138-144, 2010).

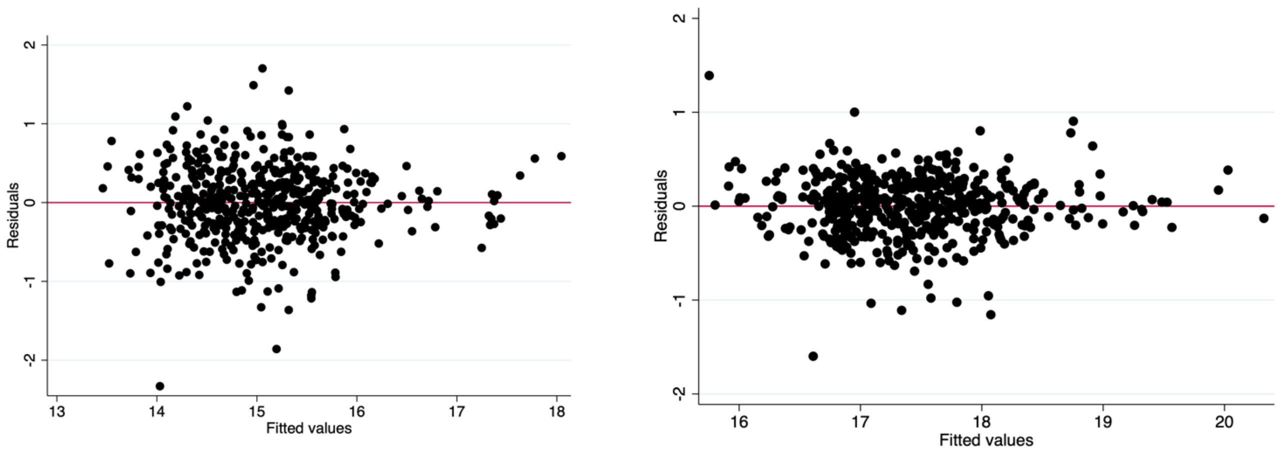

Figure II.1: Residuals vs. fitted plots (left is the footfall model and right is turnover)

The third assumption is about having no autocorrelation. Autocorrelation is about the covariance of error terms being zero over time (Brooks & Tsolacos, pp. 144, 2010). However, this study works with cross sectional data (only panel data in the robustness checks). A possible form of autocorrelation that can occur in cross-sectional data is spatial autocorrelation. The phenomenon of spatial autocorrelation has not been researched thoroughly for real estate studies. To control for spatial autocorrelation, robus standard errors are used (Ismail, 2006). When autocorrelation does occur, it only leads to problems with efficiency and not with consistency, so the coefficients will still be unbiased.

A potential problem for this research is endogeneity, meaning that the explanatory variable is correlated with the error term. Endogeneity can be suspected in the variable for vacancy. In this case, it will be related to reverse causality, where X determines Y and Y determines X. So, turnover might be determined by vacancy and vacancy by turnover. Endogeneity leads to bias in consistency and efficiency, meaning that not only the standard error, but also the coefficients are impacted. When not solving for endogeneity this may lead to untrue inferences about the coefficients. To account for endogeneity the vacancy variable is changed to the difference between the current period and the previous period. This decreased the number of observations in the dataset since there are no data available for 2016 and the previous period. However, the remaining dataset is still large enough to continue the analysis.

Finally, the fifth assumption assumes normal distribution of the error terms. When this assumption is not met, it can lead to bias in efficiency. However, when working with big samples this is often not a problem. The skewness and kurtosis test indicates that the residuals are not normally distributed. However, the dataset is large enough to have no consequences from abnormality. During the process, the variables for turnover and footfall were already transformed to the natural log to create a more normal distribution of its values.

The VIF test (table II.1 and table II.2) shows that there are no variables exceeding the VIF value of ten, so it can be assumed that there are no problems with multicollinearity. When using a more strict rule, the VIF value has to stay below five, which is not the case for the square meters of retail. However, the value is very close to five so it is assumed there is no multicollinearity problem.

Table II.1: VIF statistics model 4

Variable	VIF	1/VIF
lnsize	**5.07**	**0.197238**
lndistance	**1.85**	**0.540901**
lnvaluation	**1.13**	**0.883190**
delta_vacancy	**1.10**	**0.910508**
Center location 10,000-20,000 sqm		
Center location >100,000 sqm	**3.13**	**0.319197**
Center location 20,000-40,000 sqm	**1.99**	**0.503263**
Center location 40,000-60,000 sqm	**1.94**	**0.516197**
Center location 5,000-10,000 sqm	**2.35**	**0.426273**
Center location 60,000-100,000 sqm	**1.97**	**0.508097**
Center location <5,000 sqm	**2.85**	**0.350383**
Supportive 10,000-20,000 sqm	**1.66**	**0.602663**
Supportive 2,500-5,000 sqm	**4.87**	**0.205496**
Sopportive 5,000-10,000 sqm	**2.35**	**0.425244**
Supportive > 20,000 sqm	**4.74**	**0.211146**
Furniture mall 20,000-40,000 sqm	**1.42**	**0.702802**
Furniture mall <20,000 sqm	**2.12**	**0.471625**
Furniture mall >40,000 sqm	**1.59**	**0.628596**
Center location 10,000-20,000 sqm	**2.86**	**0.349897**
population		
<50,000	**2.89**	**0.346238**
50,000 – 100,000	**2.63**	**0.379560**
100,000 – 250,000	**2.94**	**0.340417**
250,000 +	**1.88**	**0.532249**
Mean VIF	**2.52**	

Table II.2: VIF statistics model 8

Variable	VIF	1/VIF
lnsize	**5.33**	**0.187617**
lndistance	**1.85**	**0.540901**
lnvaluation	**1.13**	**0.883190**
delta_vacancy	**1.10**	**0.910508**
SCtype		
Center location 10,000-20,000 sqm	**3.13**	**0.319197**
Center location >100,000 sqm	**1.99**	**0.503263**
Center location 20,000-40,000 sqm	**1.94**	**0.516197**
Center location 40,000-60,000 sqm	**2.53**	**0.426273**
Center location 5,000-10,000 sqm	**1.97**	**0.508097**
Center location 60,000-100,000 sqm	**2.85**	**0.350383**
Center location <5,000 sqm	**1.66**	**0.602663**
Supportive 10,000-20,000 sqm	**4.87**	**0.205496**
Supportive 2,500-5,000 sqm	**2.35**	**0.425244**
Sopportive 5,000-10,000 sqm	**4.74**	**0.211146**
Supportive > 20,000 sqm	**1.42**	**0.702802**
Furniture mall 20,000-40,000 sqm	**2.12**	**0.471625**
Furniture mall <20,000 sqm	**1.59**	**0.628596**
Furniture mall >40,000 sqm	**2.86**	**0.349897**
population		
<50,000	**2.89**	**0.346238**
50,000 – 100,000	**2.63**	**0.379560**
100,000 – 250,000	**2.94**	**0.340417**
250,000 +	**1.88**	**0.532249**
Mean VIF	**3.07**	

Table II.3: VIF statistics robustness test

Variable	VIF	1/VIF
lnsize	**9.95**	**0.100487**
lnvaluation	**2.10**	**0.475566**
vacancy_size	**1.75**	**0.572124**
SCtype		
Center location >100,000 sqm	**4.80**	**0.208173**
Center location 60,000-100,000 sqm	**2.49**	**0.401262**
Supportive from 20,000 sqm	**9.80**	**0.102041**
Furniture mall >40,000 sqm	**1.34**	**0.748839**
population		
50,000 – 100,000	**10.71**	**0.093340**
100,000 – 250,000	**14.50**	**0.068966**
250,000 +	**20.49**	**0.048804**
Mean VIF	**7.79**	

APPENDIX III: Full regression results log log

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln_footfall	ln_footfall	ln_footfall	ln_footfall	lnturnover	lnturnover	lnturnover	lnturnover
lnsize	**0.708*****			**0.588*****	**0.661*****			**0.717*****
	(19.92)			**(6.32)**	**(27.10)**			**(7.85)**
lndistance	**-0.395*****			**-0.208****	**-0.209*****			**-0.0938***
	(-5.50)			**(-3.26)**	**(-5.01)**			**(-2.44)**
delta_vacancy		**0.0108***		**-0.00321**		**0.00801**		**0.000339**
		(2.01)		**(-1.15)**		**(1.51)**		**(0.15)**
lnvaluation			**1.207**	**-0.594**			**3.280*****	**1.993*****
			(1.70)	**(-1.40)**			**(5.50)**	**(6.78)**
Center location 10,000-20,000 sqm				**0**				**0**
				(.)				**(.)**
Center location >100,000 sqm				**0.904*****				**0.303**
				(4.01)				**(1.50)**
Center location 20,000-40,000 sqm				**0.0709**				**0.0532**
				(0.78)				**(0.70)**
Center location 40,000-60,000 sqm				**0.127**				**0.0389**
				(0.91)				**(0.30)**
Center location 5,000-10,000 sqm				**-0.131**				**0.0124**
				(-1.27)				**(0.15)**
Center location 60,000-100,000 sqm				**0.402***				**0.0754**
				(2.20)				**(0.46)**
Center location <5,000 sqm				**-0.0673**				**0.0602**
				(-0.41)				**(0.51)**
Supportive 10,000-20,000 sqm				**0.113**				**0.107**
				(0.87)				**(1.28)**
Supportive 2,500-5,000 sqm				**-0.205**				**0.303***
				(-1.36)				**(2.29)**
Supportive 5,000-10,000 sqm				**-0.00223**				**0.244****
				(-0.02)				**(2.82)**
Supportive <2,500 sqm				**-0.264**				**0.407***
				(-1.38)				**(2.29)**
Supportive >100,000 sqm				**0.0705**				**0.0107**
				(0.44)				**(0.09)**
Furniture mall 20,000-40,000 sqm				**-0.912*****				**-0.497*****
				(-5.94)				**(-3.85)**
Furniture mall <20,000 sqm				**-1.085*****				**-0.625*****
				(-6.94)				**(-6.19)**
Furniture mall > 40,000 sqm				**-0.581***				**-0.276**
				(-2.52)				**(-1.42)**
Population <20,000				**0**				**0**
				(.)				**(.)**
Population 20,000-50,000				**0.0962**				**0.118**
				(1.07)				**(1.89)**
Population 50,000-100,000				**0.238***				**0.172***
				(2.28)				**(2.41)**
Population 100,000-250,000				**0.295****				**0.304*****
				(2.77)				**(4.09)**
Population 250,000+				**0.451****				**0.437*****
				(2.78)				**(4.72)**
_cons	**8.779*****	**15.06*****	**9.852****	**12.31*****	**11.49*****	**17.41*****	**3.264**	**13.13****
	(28.23)	**(387.64)**	**(3.21)**	**(6.24)**	**(53.42)**	**(515.44)**	**(1.27)**	**(2.38)**
R-squared	0.5293	0.0098	0.0062	0.7044	0.6782	0.0072	0.0608	0.8263
N	**491**	**491**	**491**	**491**	**491**	**491**	**491**	**491**

t statistics in parentheses
robust standard errors
* p<0.05, ** p<0.01, *** p<0.001

APPENDIX IV: Full regression results log level

	(1) ln_footfall	(2) ln_footfall	(3) ln_footfall	(4) ln_footfall	(5) lnturnover	(6) lnturnover	(7) lnturnover	(8) lnturnover
sqmretail	**0.0000262*****			**0.0000106*****	**0.0000249*****			**0.0000118****
	(7.69)			**(4.64)**	**(6.11)**			**(2.83)**
avdistance	**-0.127*****			**-0.0684***	**-0.0600****			**-0.0376***
	(-3.57)			**(-2.18)**	**(-2.98)**			**(-2.51)**
delta_vacancy		**0.0108***		**-0.00427**		**0.00801**		**-0.000766**
		(2.01)		**(-1.43)**		**(1.51)**		**(-0.37)**
avvaluation			**0.0165**	**-0.00614**			**0.0447*****	**0.0299*****
			(1.71)	**(-1.03)**			**(5.50)**	**(7.77)**
Population <20,000				**0**				**0**
				(.)				**(.)**
Population 20,000-50,000				**0.152**				**0.175****
				(1.67)				**(2.78)**
Population 50,000-100,000				**0.296****				**0.227****
				(2.77)				**(3.05)**
Population 100,000-250,000				**0.385*****				**0.403*****
				(3.57)				**(5.45)**
Population 250,000+				**0.463****				**0.453*****
				(2.82)				**(4.77)**
Center location 10,000-20,000 sqm				**0**				**0**
				(.)				**(.)**
Center location >100,000 sqm				**1.106*****				**0.697***
				(5.67)				**(2.21)**
Center location 20,000-40,000 sqm				**0.316*****				**0.372*****
				(3.79)				**(4.89)**
Center location 40,000-60,000 sqm				**0.532*****				**0.580*****
				(4.74)				**(4.50)**
Center location 5,000-10,000 sqm				**-0.470*****				**-0.420*****
				(-5.72)				**(-6.33)**
Center location 60,000-100,000 sqm				**0.824*****				**0.689*****
				(5.70)				**(3.37)**
Center location <5,000 sqm				**-0.765*****				**-0.826*****
				(-7.37)				**(-9.70)**
Supportive 10,000-20,000 sqm				**0.0660**				**0.0742**
				(0.51)				**(0.84)**
Supportive 2,500-5,000 sqm				**-0.836*****				**-0.499*****
				(-8.94)				**(-6.79)**
Supportive 5,000-10,000 sqm				**-0.305****				**-0.135**
				(-3.10)				**(-1.74)**
Supportive <2,500 sqm				**-1.223*****				**-0.807*****
				(-12.37)				**(-10.57)**
Supportive >100,000 sqm				**0.321***				**0.365*****
				(2.18)				**(3.40)**
Furniture mall 20,000-40,000 sqm				**-0.585*****				**-0.0307**
				(-4.10)				**(-0.26)**
Furniture mall <20,000 sqm				**-1.038*****				**-0.527*****
				(-6.01)				**(-5.01)**
Furniture mall > 40,000 sqm				**-0.166**				**0.328**
				(-0.85)				**(1.40)**
_cons	**14.88*****	**15.06*****	**13.82*****	**15.54*****	**17.13*****	**17.41*****	**14.07*****	**14.97*****
	(260.37)	**(387.64)**	**(19.14)**	**(34.87)**	**(363.21)**	**(515.44)**	**(23.21)**	**(51.80)**
R-squared	0.4059	0.0098	0.0062	0.6887	0.5290	0.0072	0.0605	0.7975
N	**491**	**491**	**491**	**491**	**491**	**491**	**491**	**491**

t statistics in parentheses
Robust standard errors
* p<0.05, ** p<0.01, *** p<0.001

APPENDIX V: Descriptive statistics robustness test

Variable	Obs	Mean	Std. dev.	Min	Max
locfootfall	**77**	**3992113**	**2244564**	**394312**	**1.13e+07**
sqmretail	**77**	**75775.32**	**61079.75**	**22000**	**284900**
avvaluation	**77**	**75.68615**	**3.486186**	**60.33333**	**82**
vacancy_size	**77**	**11.36364**	**5.527227**	**2**	**27**

APPENDIX VI: Regression results robustness test

	(11) Ln_locfootfall	(12) Ln_locfootfall	(13) Ln_footfall	(14) Ln_footfall
lnsize	**0.574*** (9.71)**	**-0.132 (-0.45)**	**0.580*** (18.27)**	**0.546*** (6.08)**
lnvaluation		**0.413 (0.28)**		**-0.540 (-1.24)**
vacancy_size		**-0.00123 (-0.10)**		**0.00121 (0.28)**
Center location 10,000-20,000 sqm		**0 (.)**		**0 (.)**
Center location >100,000 sqm				**0.786*** (3.50)**
Center location 20,000-40,000 sqm		**-0.543 (-1.60)**		**0.0699 (0.77)**
Center location 40,000-60,000 sqm				**0.102 (0.72)**
Center location 5,000-10,000 sqm				**-0.143 (-1.33)**
Center location 60,000-100,000 sqm				**0.328 (1.80)**
Center location <5,000 sqm				**-0.0941 (-0.58)**
Supportive 10,000-20,000 sqm				**0.0584 (0.46)**
Supportive 2,500-5,000 sqm				**-0.189 (-1.20)**
Supportive 5,000-10,000 sqm				**-0.00232 (-0.02)**
Supportive <2,500 sqm				**-0.247 (-1.27)**
Supportive >100,000 sqm		**-1.296* (-2.36)**		**0.0148 (0.09)**
Furniture mall 20,000-40,000 sqm				**-1.046*** (-6.73)**
Furniture mall <20,000 sqm				**-1.256*** (-8.14)**
Furniture mall > 40,000 sqm		**-1.163 (-1.91)**		**-0.699** (-2.96)**
Population <20,000				**0 (.)**
Population 20,000-50,000		**0 (.)**		**0.145 (1.46)**
Population 50,000-100,000		**0.0634 (0.25)**		**0.293** (2.60)**
Population 100,000-250,000		**0.233 (0.75)**		**0.337** (2.87)**
Population 250,000+		**0.811* (2.01)**		**0.423** (2.64)**
_cons	**8.723*** (13.23)**	**14.77 (1.83)**	**9.79*** (34.99)**	**12.36*** (6.17)**
R-squared	0.3933	0.5332	0.4737	0.6915
N	**77**	**77**	**491**	**491**

t statistics in parentheses
robust standard errors
* p<0.05, ** p<0.01, *** p<0.001

APPENDIX VII: Stata do-file

```
** Import dataset
clear all
import excel /Users/Nynke/Downloads/EKSO21

** Dropping missing values and cleaning omzetsqm
gen lnturnover =ln(omzettotaal)
drop if omzetsqm == 0
histogram omzetsqm
gen ln_omzetsqm=ln(omzetsqm)
gen lnsize =ln(sqmretail)

** Vacancy: Differences, creating new variable delta_vacancy (pasted from Excel)
*(1 variable, 970 observations pasted into data editor)

*(1 variable, 37 observations pasted into data editor)

*(1 variable, 425 observations pasted into data editor)
drop if delta_vacancy == .

** Recode benchmarkgroep
recode SCtype 8=2 2=2
label define SCtype 1 "center location 10.000-20.000 m2" 2 "Center location from 100.000 m2" 3
"Center location 20.000-40.000 m2" 4 "Center location 40.000-60.000 m2" 5 "Center location 5.000-
10.000 m2" 6 "Center location 60.000-100.000 m2" 7 "Center location op to 5.000 m2" 9 "Supportive
10.000-20.000 m2" 10 "Supportive 2.500-5.000 m2" 11 "Supportive 5.000-10.000 m2" 12 "Supportive
tot 2.500 m2" 13 "Supportive from 20.000 m2"

** Renaming variables
rename periode year
rename omzetsqm turnover_sqm
rename oppervlakte vacancy_size
rename omzettotaal turnover

** Destring location
 encode gemeente, generate(location)

 ** Added population manually for each location
. rename var19 population
// Recoding population into categories
 recode population 0/20000 = 0 20001/50000 = 1 50001/100000 = 2 100001/250000 =3 250001/max =
4
 label define Population 0 "up to 20.000" 1 "20.000 to 50.000" 2 "50.000 to 100.000" 3 "100.000 to
250.000" 4 "250.000+"
```

```
//Adding customer valuation to the dataset
** Pasted values from Excel
 rename var20 groceries
 rename var22 goaloriented
 rename var23 recreational
 egen avvaluation = rmean(groceries goaloriented recreational)

 //Adding footfall and average distance manually
** Pasted from Excel
 rename var32 footfall
 rename var33 avdistance
 drop if footfall==.

 gen ln_footfall = ln(footfall)
 gen lndistance =ln(avdistance)
 gen lnvaluation =ln(avvaluation)

//Checking the data
 scatter footfall avdistance
 scatter footfall sqmretail
 scatter footfall delta_vacancy
 scatter omzettotaal avdistance
 scatter omzettotaal sqmretail
 scatter omzettotaal delta_vacancy

 //Descriptives
 summarize footfall ln_footfall turnover lnturnover sqmretail avdistance avvaluation vacancy_size
delta_vacancy SCtype population
 corr year omzettotaal lnturnover footfall lnfootfall sqmretail avdistance SCtype population
avvaluation delta_vacancy
 tab population
 tab SCtype

 //Regression without log
 regress ln_footfall sqmretail avdistance, robust
 est store model1
 regress ln_footfall delta_vacancy, robust
 est store model2
 regress ln_footfall avvaluation, robust
 est store model3
 regress ln_footfall sqmretail avdistance avvaluation delta_vacancy i.SCtype i.population, robust
 est store model4
 regress lnturnover sqmretail avdistance, robust
 est store model5
 regress lnturnover delta_vacancy, robust
 est store model6
 regress lnturnover avvaluation, robust
 est store model7
```

```
regress lnturnover sqmretail avdistance avvaluation delta_vacancy i.SCtype i.population, robust
est store model8

esttab model1 model2 model3 model4 model5 model6 model7 model8

//Regression with log
regress ln_footfall lnsize lndistance, robust
est store model1
regress ln_footfall delta_vacancy, robust
est store model2
regress ln_footfall lnvaluation, robust
est store model3
regress ln_footfall lnsize lndistance lnvaluation delta_vacancy i.SCtype i.population, robust
est store model4
regress lnturnover lnsize lndistance, robust
est store model5
regress lnturnover delta_vacancy, robust
est store model6
regress lnturnover lnvaluation, robust
est store model7
regress lnturnover lnsize lndistance lnvaluation delta_vacancy i.SCtype i.population, robust
est store model8

//OLS tests
estat hettest
estat hettest
rvfplot
rvfplot
predict residualsff, r
sktest residualsff
predict residualsto, r
sktest residualsto
regress ln_footfall sqmretail avdistance avvaluation delta_vacancy i.SCtype i.population, robust
vif
regress lnturnover sqmretail avdistance avvaluation delta_vacancy i.SCtype i.population, robust
vif

//Heterogeneity
* Creating categories for heterogeneity tests
recode population 0/2=0 3/4=1
label define Population 0 "under 100,000" 1 "over 100,000", replace
recode SCtype 5 7 10 12 15 = 0 1 3 4 9 11 14 = 1 2 6 13 16 =2
label define SCtype 0 "small" 1 "medium" 2 "large"

*interaction regression
regress ln_footfall lnsize lndistance lnvaluation delta_vacancy i.SCtype#i.population, robust
est store model9
regress lnturnover lnsize lndistance lnvaluation delta_vacancy i.SCtype#i.population, robust
```

```
 est store model10

 esttab model9 model10

* Get descriptives
summarize ln_footfall lnturnover sqmretail avdistance avvaluation delta_vacancy i.SCtype
i.population if population<3
summarize ln_footfall lnturnover sqmretail avdistance avvaluation delta_vacancy i.SCtype
i.population if population>2

* Robustness check
//Insert Locatus data manually in old dataset with 2016 included
use "/Users/Nynke/Documents/Master/Semester 2/Master Thesis/DATA/20042022 voor droppen.dta"
rename var30 locfootfall
gen ln_locfootfall = ln(var30)
drop if ln_locfootfall==.

 regress ln_locfootfall lnsize, robust
 est store model11
 regress ln_locfootfall lnsize lnvaluation vacancy_size i.SCtype i.population, robust
 est store model12

 regress ln_footfall lnsize, robust
 est store model13
 regress ln_footfall lnsize lnvaluation vacancy_size i.SCtype i.population, robust
 est store model14

regress ln_locfootfall lnsize lnvaluation vacancy_size i.SCtype i.population, robust
vif

 //Descriptives
 sum locfootfall sqmretail vacancy_size avvaluation
 corr ln_locfootfall lnsize lnvaluation vacancy_size

//End of do-file
```