rijksuniversiteit
groningen

Demographic factors and retail vacancy

A study of Dutch districts

Tessa Mei

ABSTRACT

This thesis analyses the association between retail vacancy levels per 1,000 inhabitants and demographic factors in the districts of the Netherlands, with a particular focus on the ageing population. By conducting a multiple linear regression analysis, the findings reveal that a higher proportion of elderly residents (65+) is associated with increased retail vacancy per 1,000 inhabitants, while the youngest age group is associated with lower levels of vacancy. Additionally, areas with higher populations exhibit lower levels of vacancy per 1,000 inhabitants. Furthermore, significant rural-urban and temporal differences (before, during, and after COVID-19) were observed. The association between household income and vacancy proved insignificant.

Keywords: Real estate, retail vacancy, ageing population, demographic trends

COLOFON

Title	Demographic factors and retail vacancy: a study of Dutch districts
Version	Final
Author	T. F. (Tessa) Mei
Student number	s5139147
E-mail	t.f.mei@student.rug.nl
University	Rijksuniversiteit Groningen Faculty of Spatial Sciences Master of Science in Real Estate Studies Landleven 1, 9747 AD, Groningen
Supervisor	Dr. M. (Mark) van Duijn
Reviewer	Prof. dr. E.F. (Ed) Nozeman
Date	03/07/2024

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.

PREFACE

Over the years, I have gathered substantial knowledge, both during my Bachelor's in Human Geography and Planning in Utrecht and during my Master's in Real Estate Studies. I have utilised the acquired knowledge and academic skills in writing this thesis. My supervisor, Mark van Duijn, has been tremendously helpful with his constructive feedback, for which I am very grateful.

Additionally, I would like to extend my gratitude to the Geodienst, a service provided by the university, for answering my questions and enabling me to develop my draft maps in ArcGIS into professional ArcGIS maps. This research would not have been possible without the data from Locatus. I also want to thank them for their trust in handling the data responsibly and for providing access to it. Therefore, they will receive this thesis.

Tessa Mei

Amsterdam, June 25th, 2024

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Motivation

In November of last year, an article published in Vastgoedmarkt captured attention with its headline: "Ageing and the retail sector: threat or opportunity?" (Schröder, 2023). The piece delves into projections indicating a potential 2% revenue loss in the retail sector by 2032 due to the ageing population. Such concerns were echoed earlier, in October 2022, by ABN Amro, highlighting the evolving landscape of the retail sector in response to population demographics (ABN AMRO, 2022). According to ABN AMRO's analysis, the ageing population tends to curtail spending in non-food retail segments, as older individuals typically possess more items, spend less, and allocate more time to leisure activities rather than shopping (ABN AMRO, 2022).

As demographic shifts unfold, with one in four Dutch citizens projected to be 65 years or older in nine years' time, the retail market faces a pressing need to adapt (ABN AMRO, 2022). However, amidst these demographic changes, Locatus reported a noteworthy development: retail vacancy in the Netherlands has decreased to its lowest level in over a decade (Custers, 2023). Given the acknowledged impact of retail vacancy on the liveability of city centres, this reduction is a significant development (Lagerwaard, 2022).

Against this backdrop of demographic shifts and evolving retail vacancy trends, this research endeavours to explore the relationship between demographic structures within Dutch cities and retail vacancy levels. Specifically, the analysis aims to discern whether variations in age demographics have influenced retail vacancy since 2019 and, if so, whether this influence has evolved in recent years. The selection of 2019 as the starting point for the analysis is deliberate, as it marks the commencement of a new vacancy cycle, as depicted in Figure 1.

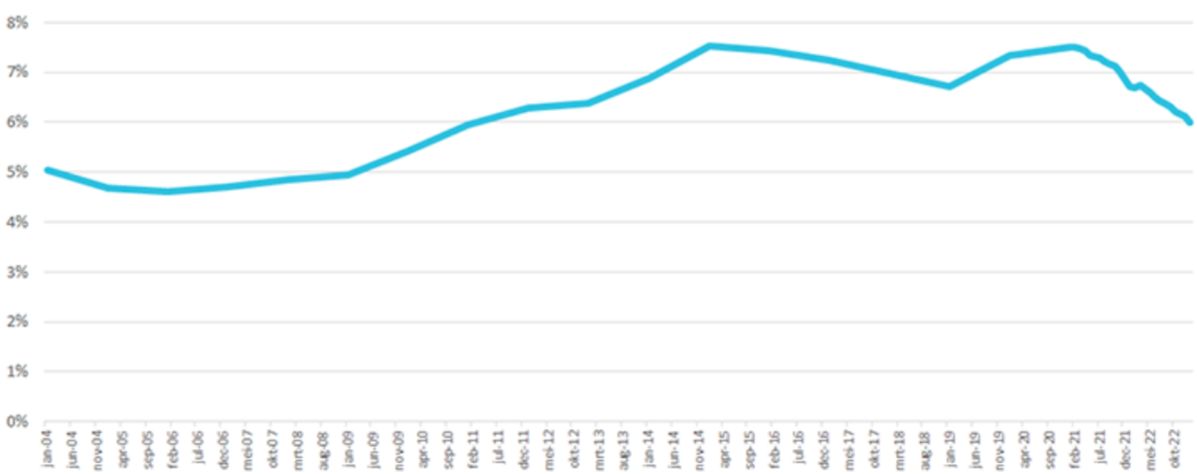

Figure 1: Development retail vacancy in the Netherlands in % (Slob, 2023)

The commencement of the new vacancy cycle in 2019 coincides with the onset of the COVID-19 pandemic, a period during which a significant surge in vacancy might have been expected (PBL, 2022). However, contrary to expectations, such a surge did not materialise. Despite the challenges encountered by the Dutch retail sector during the pandemic, including lockdowns and shifts in consumer behaviour, governmental intervention through financial support measures aimed to bolster businesses and prevent bankruptcies. These measures played a crucial role in maintaining the stability of the retail sector and mitigating the impact on retail vacancy to just a few percentage points.

Given these conditions, it is important to investigate whether regional differences exist in how the pandemic has affected retail vacancy. These differences may stemfrom variations in local economic strength as well as demographic profiles. A thorough understanding of these complexities could help identify effective policies to support the retail sector in the future.

1.2 Academic relevance

Several studies have been conducted on retail vacancy, offering various causes for this phenomenon. Benjamin, Jud, and Winkler (1998) posited that retail vacancy arises from the disparity between supply and demand, where insufficient demand relative to supply leads to vacant spaces. Mallach's (2018) research aligns with this perspective. He investigated vacancy and "hyper vacancy" in the United States, contending that vacancy serves as a symptom of other issues such as market failure, concentrated poverty, and economic decline. Dolega and Lord (2020) explored the retail market in Liverpool and asserted that vacancy is also contingent on the location of the store. Furthermore, Talen and Park (2021) in their qualitative study, suggested that demographic shifts are frequently cited explanations for retail vacancy, stating that the relationship between demographic changes and retail vacancy is "in need of study", lending valuable scientific support to this research.

Delage et al. (2020) put forth "demographic dynamics" as a reason for the retail decline in French cities, defining these dynamics as growing or falling populations. Campo et al. (2000) argued that retail sales are based on local market potential and purchasing power, influenced in turn by population characteristics such as family size, income level, ethnicity, and age distribution. Another study on retail sales is that of Meltzer and Capperis (2016), focusing on New York City, asserting that population growth predicts higher sales. Larger households and a higher proportion of white residents are associated with lower sales.

The studies by Delage et al. (2020) and Meltzer and Capperis (2016) suggest that demographic changes, including population growth and age distribution, may influence retail vacancy but do not explicitly address this relationship. This highlights a gap in the literature. Existing scholarly literature on retail vacancy presents various causes, yet there is a notable absence of quantitative research comprehensively exploring the relationship between demographic changes, including the ageing population, and retail vacancy in the Netherlands. The qualitative approach of Talen and Park (2021) provides a scholarly rationale for further investigation. Consequently, this research proves to be a valuable addition to the existing literature, contributing to a deeper understanding of retail vacancy phenomena. Moreover, it can offer valuable insights for policymakers and stakeholders in the retail sector to develop and implement more effective strategies for urban development and economic vitality.

1.3 Research problem statement

This research aims to examine the relationship between demographic levels and retail vacancy in the Netherlands, with specific attention given to the ageing population.

The societal and scientific relevance has led to the formulation of the following main research question:

“To what extent are demographic factors associated with retail vacancy in the districts of the Netherlands?”

To address the main research question, several sub questions need to be answered. The sub questions are:

1. What are the drivers of retail vacancy according to academic literature?
2. To what extent do levels of population size, age distribution, and household income associate with retail vacancy in the districts of the Netherlands between 2019 and 2023?
3. To what extent do population size, age distribution and household income contribute to retail vacancy disparities between urban and rural districts in the Netherlands?
4. To what extent do population size, age distribution and household income contribute to differences in retail vacancy before, during and after COVID-19?

Retail vacancy, however, can be influenced by multiple variables, implying that the relationship between dependent and independent variables is considered probabilistic. Furthermore, control variables are utilised, which are kept constant during the study to enhance the quality of the results. The control variables consist of “retail property characteristics”(urbanity level, year of construction, and store size) and are incorporated into the conceptual model in Figure 1. Urbanity level means the area type (urban or rural) where the district is located. The store size is measured through the retail selling floor area. Figure 1 illustrates the relationship between the independent variables and the dependent variable.

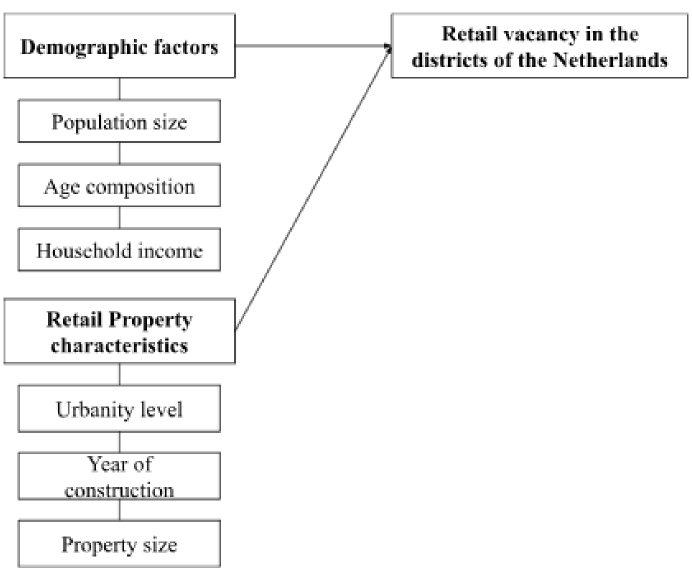

Figure 2: Conceptual model (own edit)

1.4 Outline

The rest of this paper is organised as follows. In Chapter 2, additional clarification is provided regarding the existing theory, accompanied by a literature review on the factors and causes contributing to retail vacancy. Additionally, this section addresses the first sub question. Chapter 3 delves into a more detailed explanation of the data collection and methodologies employed. Chapter 4 outlines the results and answers sub question 2, 3 and 4. This section also engages in a discussion grounded in the existing academic literature. Chapter 5 concludes based on both theoretical and empirical findings, addressing the primary research question. Moreover, this chapter explores policy implications, the study's limitations and proposes avenues for future research.

2. THEORY, LITERATURE REVIEW AND HYPOTHESES

2.1 Vacancy

Vacancy is defined in multiple ways in different studies. Van der Voordt et al. (2007) describe vacancy as the situation where a space offered for rent is not currently leased. Rabianski (2002) offers a more detailed definition from a supply perspective, defining vacancy as the presence of excess supply at a given price in the property market. Similarly, Benjamin et al. (1998) define vacancy as a mismatch between the demand and supply of retail space in terms of amount and location.

Vacancy is frequently categorised based on the duration of the unoccupied period (Van der Voordt et al., 2007; Evers et al., 2014). When a retail space remains vacant for less than a year, it falls into the classification of "frictional vacancy." This form of vacancy does not give rise to significant issues and is a consequence of the time required to identify new lessees. The frictional vacancy is, to a certain extent, imperative for the smooth operation of the economy, facilitating the expeditious identification of new business entities (Evers et al., 2014). Vacancy persisting between one and three years is denoted as "prolonged," while vacancy exceeding three years is labelled as "structural" (Van Gool et al., 2007). Prolonged and structural vacancies contribute to an aesthetically unpleasing urban landscape and diminish the overall appeal of the commercial thoroughfare, necessitating potential interventions by municipal authorities (Buitelaar et al., 2013).

<u>The Four-Quadrant Model</u>

This model can give an explanation for vacancy in the property market, in this case, the retail property market. DiPasquale and Wheaton's (1992) Four-Quadrant Model offers insights into the market dynamics for real estate space and assets. The model in Figure 3 comprises the northeast (NE) quadrant representing the property market, the northwest (NW) quadrant for capitalisation rates, the southwest (SW) quadrant for construction costs, and the southeast (SE) quadrant for the impact of market movements on the stock of real estate space.

For the retail market (NE quadrant), equilibrium is achieved when demand equals the stock of space, influenced by factors such as retail rents, sales, vacancy, development constraints and demographic factors such as population dynamics local income and age composition. The supply of retail space is determined by developers, influenced by the economic climate, land availability, capital market cycles, interest rates, and local conditions, leading to imbalances and resulting in retail vacancy.

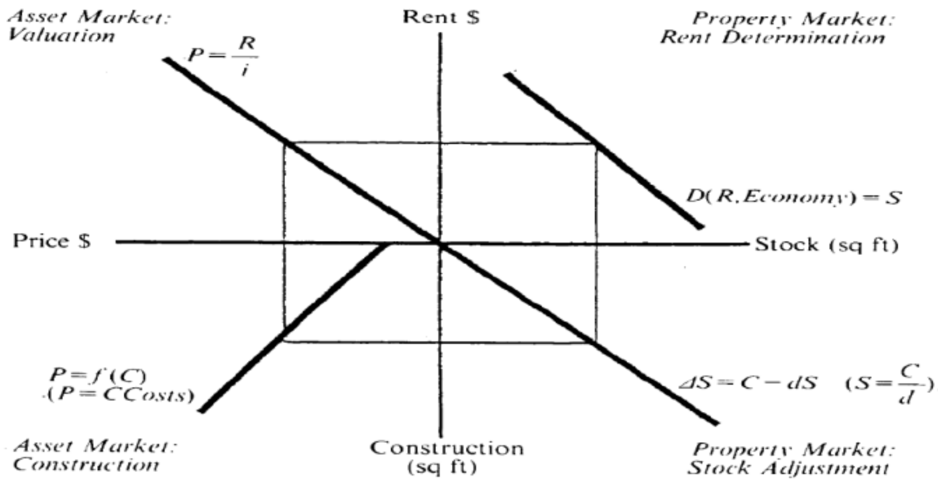

Figure 3: Four-Quadrant Model by DiPasquale and Wheaton (1992)

However, vacancy is not explicitly incorporated into this model by DiPasquale and Wheaton (1992). This is because the underlying assumption of their model is that the price will decrease due to excess supply, diminishing the attractiveness of new real estate development while simultaneously fostering an increase in demand. As a reaction to this, the model predicts that a decline in supply will follow, and therefore, the market will autonomously reestablish an equilibrium. However, this scenario does not align with reality; otherwise, there would be no occurrence of (prolonged and structural) vacancy.

In contrast, Colwell (2002) extends the Four-Quadrant Model by integrating vacancy dynamics, introducing the concept of natural vacancy. Natural vacancy includes frictional vacancy as described in 2.1, as well as other factors that contribute to the overall level of vacancy in a healthy market, such as seasonal trends or market cycles. So this concept accounts for tenant turnover and the time required to fill vacant units (Read, 1988). Colwell integrates natural vacancy into the SE quadrant, to utilise this in the NE quadrant. Figure 4 illustrates how natural vacancy, denoted as vS1, extends from the current stock's vacancy function to the horizontal axis at a 45-degree angle. This enhancement by Colwell (2002) contributes significantly to a more comprehensive understanding of vacancy dynamics within the Four-Quadrant Model.

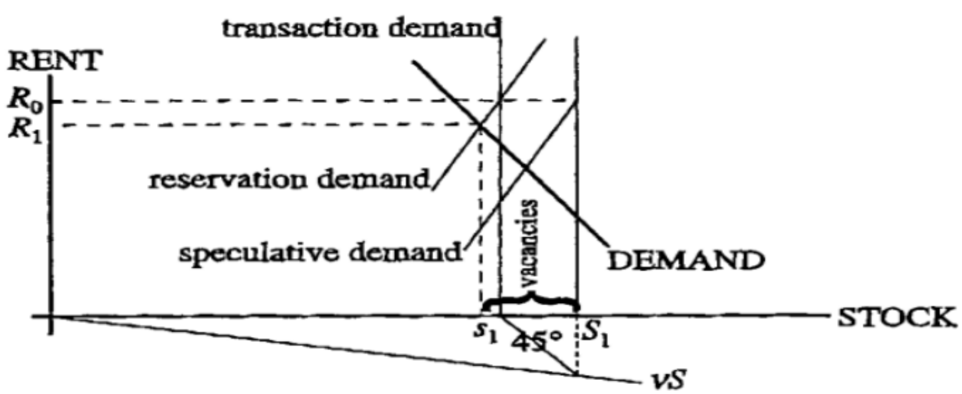

Figure 4: Addition of vacancy to the Four-Quadrant Model (Colwell, 2002)

2.2 Demographic determinants of vacancy

Several studies have been conducted on the causes of retail vacancy. Reasons such as the increasing popularity of E-commerce and factors related to the shopping area itself are often mentioned (Brouwer & Tool, 2018). In this theoretical framework, the drivers for retail vacancy will be based on the independent (demographic) variables of this study. Theories explaining the association with retail vacancy will be formulated for each determinant.

Age composition

The shopping behaviour of older individuals differs in several aspects from that of younger people. For instance, older individuals tend to exhibit more loyalty towards stores, expect personalised attention, take longer to make decisions in a store, and view shopping as a social event (Brenner and Clarke, 2017; Lesáková, 2016). The latter is particularly interesting as it alludes to the social aspect of shopping. To better understand this phenomenon, a "shopping as practice" approach is adopted, drawing on the social practice theory (Hansson, Holmberg, and Post, 2022). This theory contributes to a better understanding of routines in daily life, such as shopping. The practice is conceptualised as an entity consisting of three main aspects: materials, meanings, and skills. Each of these three aspects is necessary for performing the practice, in this case, shopping. The model implies interdependence among the three aspects, suggesting that any shift in one of them will impact the overall practice (Reckwitz, 2002; Shove, Pantzar & Watson, 2012; Warde, 2005; Warde, 2014). Figure 5 illustrates the model of this theory.

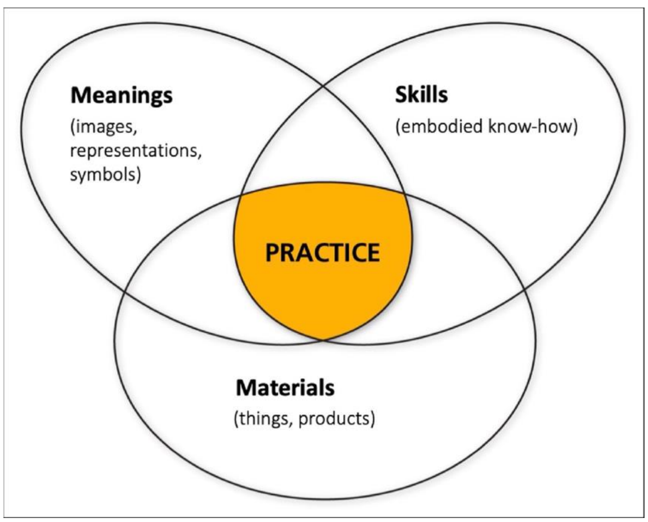

Figure 5: Social practice theory model (Morgan et al., 2022)

The meanings in this theory pertain to the social aspect of shopping. The materials on the retail products and the skills relate to shopping skills. Compared with the social practice of online shopping, where skills may be lacking for older individuals, this will negatively impact the social practice of online shopping for seniors. This theory also confirms that the social aspect of shopping for older individuals has an effect on the shopping practice, essentially influencing shopping for seniors.

Contrary to the preceding discussion, research indicates a projected 2 per cent revenue decline for the retail sector in the Netherlands due to the increasing ageing population, as highlighted in Chapter 1.1. This phenomenon is closely tied to the consumption and spending patterns of various age groups. Households with a senior citizen aged 65 and above as the primary breadwinner tend to exhibit lower consumption levels compared to "younger" households in the Netherlands (CBS, 2017). However, they also possess lower disposable incomes. Interestingly, despite their lower income levels, older households tend to allocate a relatively larger portion of their earnings toward consumption than their younger counterparts. A significant proportion of the expenditures made by older households is directed towards home furnishings, charitable contributions, and healthcare. Thus, while seniors earn less than individuals in other age brackets, they allocate a greater share of their income to consumption, potentially impacting the effect of ageing on retail vacancy rates in both reinforcing and attenuating manners.

Population size

A decrease in population size may be associated with an increase in retail vacancy (Hollander et al., 2018). A scientific theory linked to this is the "Retail Gravitation Theory" (also known as Reilly's Law of Retail Gravitation). This theory focuses on predicting the attractiveness of shopping centres to consumers based on the distance to the centre and the size of the population (Reilly, 1931; Brown, 1992). According to this theory, consumers are likelier to go to the nearest shopping centre unless another centre has a significantly larger catchment area. The population size and the distance to alternative shopping centres influence the attractiveness of a shopping area. The Retail Gravitation Theory implies that the population size in a region influences the attractiveness of different shopping centres, which may contribute to the emergence or reduction of retail vacancy over time.

The "Economic Base Theory" may also relate to the relationship between population size and retail vacancy. This theory focuses on the economic structure of a region and the relationship between basic and non-basic sectors. The theory posits two main sectors in an economy: the basic sector and the non-basic sector. The basic sector includes activities targeted at external markets, such as exports, which are crucial for the region's economy. The non-basic sector comprises activities primarily focused on the local market. Population size impacts the basic sector in a region (Wang and Hofe, 2008). If the population size increases and there is a strong basic sector leading to more employment and income, this can result in a higher demand for local retail amenities. Conversely, if the economic base is weak with little external activity, this may lead to reduced demand for local shops and potential retail vacancies.

Household income

Demographic shifts, particularly changes in income distribution, exert a notable influence on retail vacancy rates, as posited by Bieniek et al. (2018) and Kickert (2021) in alignment with Keynes' Income-Expenditure Model. Their argument underscores the significance of income disparities between high and low-income segments of the population, which manifest in distinct consumption patterns across specific geographical areas. This assertion is complemented by Mallach's (2018) observation that concentrated poverty, delineated by household income levels, serves as a potential catalyst for increased vacancy rates in retail spaces. Mallach's perspective, previously highlighted in Section 2 of Chapter 1, underscores the adverse impact of concentrated poverty on retail environments. Additionally, Kang (2019) contributes to this discourse by examining the relationship between higher income levels and augmented retail turnover. Kang's findings emphasise that affluent demographics, characterised by increased spending propensity, drive heightened retail activity. Notably, Kang (2019) underscores the significant role of neighbourhood income levels over consumption patterns in influencing retail turnover dynamics, further enriching the understanding of the complex interplay between income composition and retail outcomes.

2.3 Spatial determinants of vacancy

To address research question 3 comprehensively, a deeper examination of location theories is necessary to investigate the disparities in retail vacancy between rural and urban areas, which are called spatial determinants of vacancy.

A theory that provides insights into what drives actors in choosing locations is the Bid-Rent Theory. This theory fundamentally builds upon Von Thünen's Concentric Zone Theory. The Bid-Rent Theory posits the existence of a centre of economic activity, where each actor aims to establish themselves as closely as possible, minimising transportation costs (Alonso, 1964; Adhvaryu, 2010). In this model, this centre of economic activity is referred to as the "Central Business District" or CBD.

Figure 6 illustrates the impact of distance from the CBD on rent. Retail is situated closest to the CBD, as retailers (apparently) can afford higher housing costs than other bidders and benefit the most from a central location. This is where the highest concentration of consumers occurs, maximising the likelihood of high turnover or profit. According to the Bid-Rent Theory, offices and industrial areas are positioned somewhat farther from the centre, with residential areas extending beyond. This distribution is approximate, as there are instances of shops located in residential zones and vice versa. Nevertheless, the model continues to provide a robust explanation for the location choices of parties in contemporary settings (Reimers & Clulow, 2004).

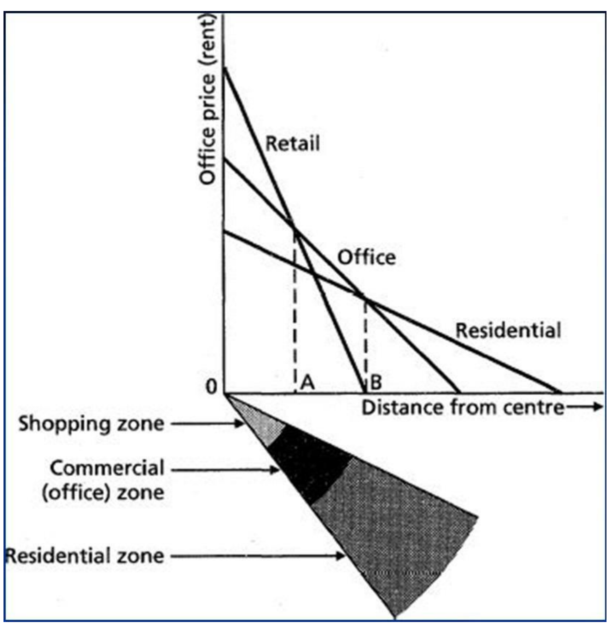

Figure 6: Bid-Rent Theory (Adhvaryu, 2010)

In the “Central Place Theory”, Christaller examines an optimal localisation of shopping areas. The theory posits that there should be a sufficient distance between different centres for the placement of specific shopping centres (Peek & Veghel, 2011). This can be seen in Figure 7. The model is highly abstract and assumes an equal distribution of people across space and distinguishes between main and sub-retail locations.

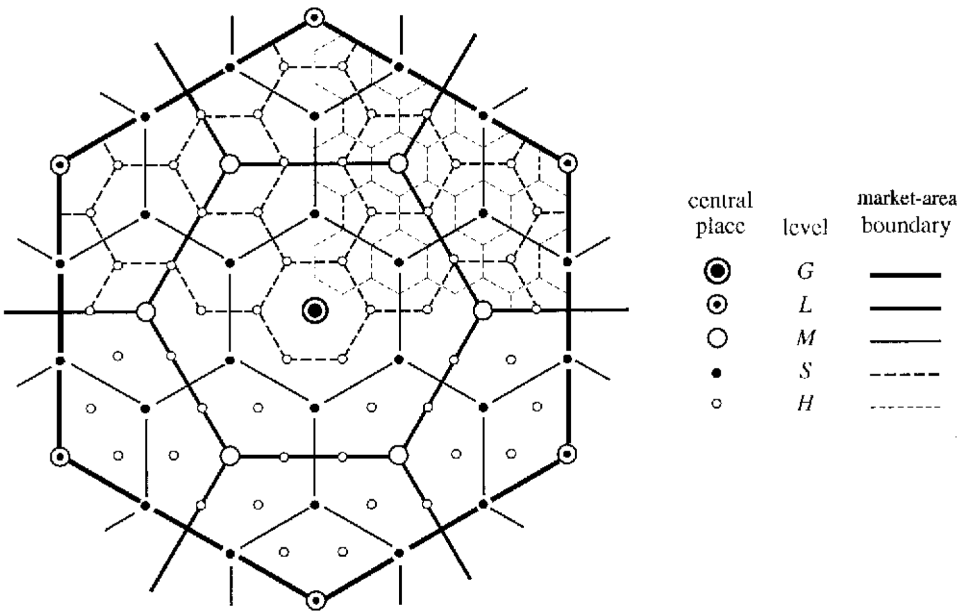

Figure 7: Central Place Theory (Ishikawa and Toda, 2000)

Christaller's model continues to be observable in urban centres as shopping areas, district shopping centres, and neighbourhood shopping centres, each with its own catchment area (Peek & Veghel, 2011). This spatial retail policy has directly emerged from Christaller's theory. A basic formula of this theory can be represented as: $Ai = f(P, \ldots)$. Here, Ai represents the market value of location i and P represents a population variable.

However, Christaller does not address vacancy but emphasises hierarchy. Changes in consumer behaviour, such as a greater preference for stores in the main core (or, conversely, outside it), can occur. Increased automobile accessibility is an example of this shift. Furthermore, the uneven distribution of income leads to a practical scenario that significantly deviates from the Christaller model. The resistance to reaching a location also varies per shopping centre, with accessibility crucial in determining a centre's popularity (Bolt, 2003).

The hypothesis derived from the aforementioned theories regarding vacancy in rural and urban areas is straightforward. In interpreting the CBD as an urban area and considering the Bid–Rent Theory, where the greater the distance from the CBD, the more rural it becomes, it can be inferred that the further a store is located from the CBD, the higher the likelihood of vacancy due to factors such as the decreasing concentration of consumers and, consequently, the store's profitability. The concentration of consumers can be linked to the demographic factor of population growth, while the profitability of a store can be associated with household income, as discussed in Chapter 2.2.

Christaller's theory underscores the significance of location and distance in organising stores and determining their catchment areas. To draw conclusions about the relationship between demographic factors and retail vacancy in rural and urban areas, population growth is considered, along with speculation on what Christaller would anticipate regarding subsequent changes in catchment areas. If a main retail location, in Christaller's theory, serves as a central place (urban area) and experiences sudden population growth, it could have various effects on the local spatial system. Firstly, population growth may increase the demand for goods and services, potentially expanding the city's market area (catchment area) beyond initial projections. This could lead to an expansion of the city's threshold, enabling it to offer a wider range of specialised goods and services than initially expected. Consequently, this might result in the further clustering of retail in the central place (urban area) and increased vacancy in the catchment area (rural area). Conversely, if population growth occurs in a sub-retail location considered rural, this could potentially lead to vacancy in main retail locations as the catchment area for these main retail locations diminishes.

A sequel to Christaller's work and an expansion of the Retail Gravitation Theory Model, as discussed in Chapter 2.2, has emerged to fortify Christaller's theory by incorporating considerations of population growth. This extension of the gravity model by Harris in 1950 enables the integration of the demographic factor of household income to explore vacancy rates within the framework of the central place theory. Here, the population variable (P in Christaller's formulation) influencing Ai (the market value of location i) is contingent upon various factors, including aggregate disposable income. Consequently, a rise in income levels prompts an augmentation in P, thereby fostering an increase in Ai. In instances where Ai exhibits substantial market value, minimal vacancy is anticipated, while areas characterised by lower market values may experience elevated vacancy rates (Thrift and Kitchin, 2010).

2.4 Other determinants of vacancy

Remøy et al. (2007) identified characteristics of office buildings, such as the construction year, size, status, height, and property price, as well as location characteristics like car accessibility, parking facilities, and proximity to the city centre, that influence office vacancy in the Amsterdam region. These findings could potentially be relevant for retail properties as well. Hoekstra and Vakili-Zad (2011) support the conclusions of Remøy et al. (2007), asserting that the construction year affects vacancy in Spain. They argue that older buildings have a higher tendency for vacancy compared to newer ones. Yakubu et al. (2017) also confirm that older buildings exhibit higher vacancy rates. Furthermore, Ball (2002) emphasises that the size of a building determines vacancy in the United Kingdom; smaller buildings are more frequently vacant and are more suitable for redevelopment. The research by Remøy et al. (2007) is further reinforced by the proposition that in more urban environments, there is a greater potential for vacancy and transformations of properties. Oskam (2021) supports this finding by stating that retail spaces are less frequently vacant in areas with lower levels of urbanity than areas with higher levels of urbanity. Furthermore, online shopping can determine vacancy levels (Zhang, Zhu and Ye, 2016). Online shopping is something that even older age groups have learned due to COVID-19, which could influence shopping behaviour in the after COVID-19 period (Ministerie van Infrastructuur en Waterstaat, 2022).

2.5 Hypotheses

The hypotheses have been formulated based on the evaluation of existing literature and underlying theories. This research examines the relationship between demographic changes and retail vacancy, focusing on age composition, population size and household income, which are the independent variables discussed in Section 2 of this chapter. The independent variables and their expected relationship to the dependent variable, retail vacancy, are presented in Table 1.

Table 1: Hypotheses independent variables

Independent Variable	*Expected sign*
Age composition	-
Population size	-
Household income	-

H1: The age composition, population size and household income are expected to be negatively associated with retail vacancy in the districts of the Netherlands.

The hypothesis that arise based on sub question 3 and 4 are presented below.

H2: The demographic factors will contribute to retail vacancy disparities between rural and urban districts in the Netherlands.

H3: The demographic factors will contribute to differences in retail vacancy before, during and after COVID-19.

3. DATA AND METHODS

3.1 Context

Figure 8 depicts the evolution of vacant retail premises in the Netherlands. A noticeable trend is evident, showing a gradual decline in the number of physical retail stores alongside a rise in the number of online retailers. In 2019, the Netherlands counted 85,921 physical retail stores, which decreased to 82,124 by 2023.

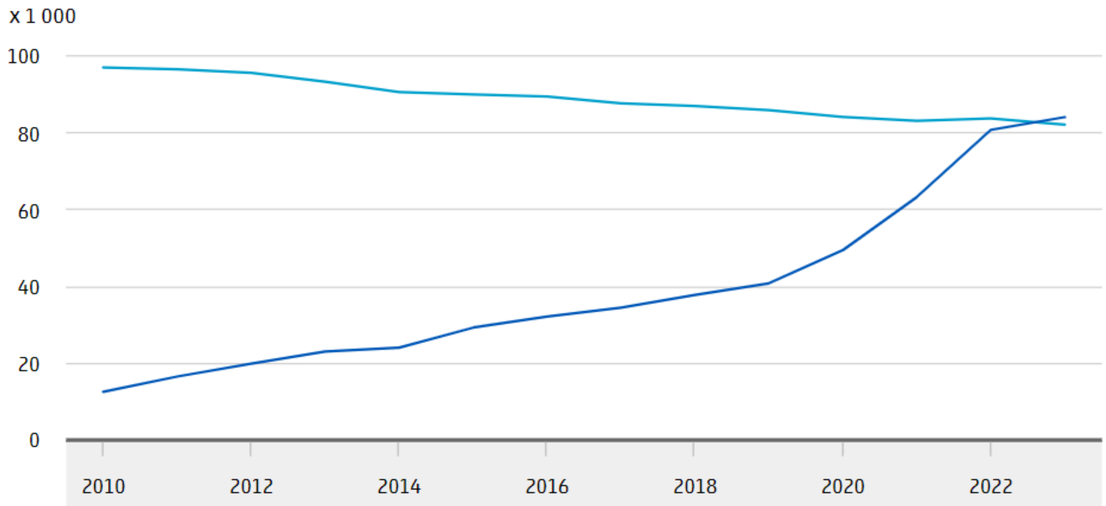

Figure 8: Total amount of retail properties on the 1st of January (CBS, 2023)
Note: The light blue line represents physical retail stores, while the dark blue line represents online stores.

The annual count of vacant retail premises does not correspond with the declining trend depicted in Figure 8. Figure 9 presents the total number of vacant retail premises per year in the Netherlands, inclusive of those that remained vacant from the previous year. The table reveals a minor peak in vacancy rates in 2020 and 2021, followed by a decline to the lowest count in this series by 2023. This trend is largely attributed to the COVID-19 pandemic situation prevalent during that period, as described in Chapter 1.

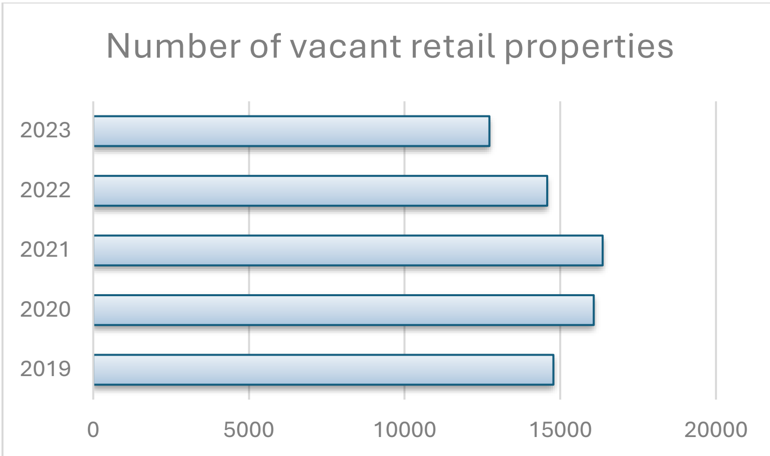

Figure 9: total amount of vacant retail properties per year in the Netherlands (Locatus, 2023)

3.2 Data collection and operationalisation of variables

The datasets utilised to conduct the regressions in this study originate from two distinct sources. One source is Locatus, which supplied data for the dependent variable, while the other one is Statistics Netherlands (CBS), which provided data for the independent variables. This division of data procurement underscores a deliberate selection process aimed at utilising comprehensive datasets from reputable sources to facilitate robust regression analyses.

Locatus

Locatus is an organisation specialised in the collection, analysis, and provision of detailed information pertaining to the retail sector. They maintain an extensive database containing data on stores, shopping centres, and other commercial premises. This information is utilised by various stakeholders for market research, location analysis, and strategic decision-making. However, access to Locatus datasets is not available through their website; rather, it is only accessible upon request. For this research, a request was made for five databases concerning retail vacancies, which were accepted and subsequently retrieved from their headquarters in Utrecht. An employee logged into the Locatus database portal on the computer to retrieve the datasets, as the passwords were not permitted to be known. After logging in, only the agreed-upon five datasets were downloaded from their portal under the employee's supervision. These five Excel files were saved to the hard drive on my laptop. Upon logging out of the portal, it was verbally agreed that this research would be shared with them once it is completed. A non-disclosure agreement was deemed unnecessary.

The obtained datasets encompass information on all vacant premises from the years 2019 to 2023, because 2019 marked the commencement of a new vacancy cycle, as described in Chapter 1. Within the scope of this study, this data facilitates the identification of the location of vacant premises per year (measured by Locatus once per year), thereby enabling the determination of vacancy levels per district for use as the dependent variable. So the analysis focuses solely on information pertaining to vacant retail premises rather than encompassing all retail establishments, both vacant and non-vacant. Consequently, the vacancy level is quantified instead of the vacancy percentage. However, this is only one level and does not provide enough insight in analyses because it represents an absolute number. As a result, the hypotheses would not be accurate, since an increase in population would lead to more vacancy due to more people, thus more stores, and consequently more vacancy. Therefore, district size was controlled for. The new dependent variable is the amount of vacancy per 1,000 inhabitants. Performing the regressions revealed that this approach optimises the results compared to using the absolute number of vacant properties per district as the dependent variable. Additionally, changes in vacancy per district were considered as the dependent variable, but this resulted in too many lost observations.

Locatus employs the following definition of vacancy: "A premise is registered as vacant if there is a reasonable expectation that a retail outlet, catering establishment, or consumer-oriented service will be established therein." Additionally, they categorise vacancy into three categories: initial and frictional vacancy (less than 1 year), prolonged vacancy (1-3 years), and structural vacancy (more than 3 years). This study does not differentiate between types of vacancy, as the focus lies more on the quantity of vacancy per district per 1,000 inhabitants in comparison with demographic factors rather than specifically the type of vacancy per district in relation to demographic factors. Another rationale is that the distribution between the three categories has remained relatively stable during the 2019-2023 period (Locatus, 2023).

Statistics Netherlands (CBS)

The Central Bureau of Statistics (CBS) is the national statistical organisation of the Netherlands. It collects, processes, analyses, and publishes a wide range of statistical data about Dutch society. The CBS gathers data through surveys, administrative records, and other sources, generating statistics on various topics, including population, economy, employment, income, health, education, crime, and the environment. The extensive datasets containing these statistical data are publicly accessible online. The datasets utilised for this research consist of five "Neighbourhood and District Key Figures" (KWB) datasets, covering the years 2019-2023. These datasets provide district-level data on the independent variables of this study, including information on population size per district, the distribution of population across different age groups, and data on average standardised income. In addition to these KWB datasets, the CBS shapefiles of neighbourhoods and districts per year were employed to obtain district-level data from Locatus in ArcGIS Pro, as previously mentioned.

The obtained datasets from Locatus contain information regarding vacant premises and their respective locations. Initially, this information was available at the municipal and postal code levels, as well as at the x-y coordinate level, but not at the district level. Given that this research focuses on district-level analysis, a conversion of this data to district level was necessitated using the x and y coordinates. This conversion was performed in ArcGIS Pro utilising district shapefiles provided by the CBS. By executing a Spatial Join operation in the geoprocessing toolbox, the district names and codes were appended to the vacant premises from the Locatus databases. Subsequently, these newly created attribute tables for each year were imported into Excel, where they could be merged with the "Neighbourhood and District Key Figures" (KWB) datasets for each respective year based on district codes. After merging all years with the CBS data, efforts were made to ensure consistency across all five merged datasets, enabling the combination of all five years into a single Excel sheet. To achieve this, redundant columns were removed. The final Excel sheet encompassing all years contains information on all districts of the Netherlands for each year, with vacant premises data appended if any premises are vacant within the respective districts. In cases where multiple premises are vacant within the same district, the district appears multiple times in the sheet. This approach was adopted based on advice from the Geodienst, ensuring that individual data regarding vacant premises is retained. However, this caused issues with the regressions in STATA later on due to multiple instances of the same districts per year, resulting in too many duplicates. Therefore, these duplicates were removed in STATA in a later stage. Districts wherein no premises are vacant, are also included in the sheet for each year; however, the corresponding columns pertaining to vacant premises information remain empty for the Locatus data. This Excel sheet was then imported into STATA, where a variable titled “Vacancy across all years” was created to compute the quantity of vacant premises per district for each year. This variable was created prior to removing duplicates of neighbourhoods per year, as otherwise the number of vacant properties per neighbourhood could not be calculated. By dividing this variable by the population size and multiplying by 1,000, the dependent variable was generated.

The independent variable “population size” is assessed by the number of inhabitants per district from the CBS KWB dataset. This entails the count of residents per district. “Household income” is measured by the average standardised income per district multiplied by 1000 euros. This metric represents disposable income adjusted for variations in household size and composition. This adjustment is conducted using equivalence factors, which capture the economies of scale resulting from shared household living. Through equivalence factors, all incomes are standardised to that of a single-person household, rendering household welfare levels comparable. Standardised income serves as a measure of household (members') prosperity. However, data pertaining to standardised income is lacking only for the year 2023, as it has not yet been collected by the CBS. “Age composition” is evaluated by examining the number of inhabitants falling within specific age categories. These categories are delineated in the KWB dataset as follows:

- 0 to 15 years. Number of inhabitants aged 0 to 15 years on January 1st
- 15 to 25 years. Number of inhabitants aged 15 to 25 years on January 1st
- 25 to 45 years. Number of inhabitants aged 25 to 45 years on January 1st
- 45 to 65 years. Number of inhabitants aged 45 to 65 years on January 1st
- 65 years or older. Number of inhabitants aged 65 or older on January 1st

Because this independent variable is closely related to the population size, multicollinearity issues may arise later when conducting regressions in STATA between these variables. To mitigate this, the population count per age group has been transformed into the share of residents per age group. For this purpose, a variable was created containing the total population per neighbourhood per year, allowing the population counts per age group to be divided by this total and multiplied by 100.

The control variable “property size” is measured using data from the Locatus datasets pertaining to the WVO (Winkelverkoop Vloer Oppervlakte). Locatus defines this as "The area of a unit that is freely accessible or visible to the public, including spaces directly associated with sales and/or services." Furthermore, Locatus only measures WVOs for premises exceeding 100 square meters. Smaller spaces are visually estimated (Locatus, 2023).

The control variable “Year of construction” is also assessed using data from the Locatus datasets, specifically the “BAG Bouwjaar” (Building Year). This refers to the original construction year of the vacant retail premises. Locatus obtained this data from the Basisregistratie Adressen en Gebouwen (BAG)

Netherlands. However, Locatus only began incorporating this variable into their datasets in 2020. This implies that the data for the year 2019 is missing.

The control variable “Urbanity level” pertains to the third sub question of this study, focusing on the disparity in vacancy between rural and urban areas. For this, the environmental address density (OAD) per square kilometre from the CBS is utilised. The OAD of a neighbourhood, district, or municipality is the average number of addresses per square kilometre within a circle with a radius of one kilometre on January 1st of the respective year. The OAD aims to depict the degree of concentration of human activities (residential, employment, education, shopping, recreation, etc.). The CBS employs the OAD to determine the urbanisation level of a specific area. In this study, a dummy variable is created for this variable. The CBS itself has classified the degree of urbanisation based on the OAD into 5 categories, ranging from very strongly urban (>2500 addresses per km2) to non-urban (<500 addresses per km2). In this study, it is decided that areas with up to 1000 addresses per km2 are categorised as rural, and those with more than 1000 addresses per km2 are urban. Rural is denoted as 0, and urban as 1. This decision is based on CBS's definition of rural areas: "If the OAD is less than 1000 addresses per square kilometre, it is classified as rural."

3.3 Descriptive statistics

In accordance with the research question delineated in Chapter 1, this study investigates the interplay between the dependent variable, namely the level of retail vacancy per 1,000 inhabitants, and various district characteristics, including population size, age distribution, and household income. The continuous variable "number of vacant retail units per 1,000 inhabitants" is adopted as the dependent variable, coupled with retail-related information, to underpin the investigation. As discussed in the previous paragraph, the Locatus data at the neighbourhood level was obtained by layering it with a shapefile in ArcGIS. These Excel datasets from Locatus at the district level were then transferred to MS Access, where each year's Locatus dataset was merged with the corresponding KWB dataset based on district codes. These years were then combined into one Excel file and imported into STATA. Here, redundant variables were removed, dummy variables were created, averages were taken of variables such as construction year and WVO (weighted floor area), and duplicates were dropped. This extensive data cleaning process led to the original number of observations of 78,555 decreasing to 15,994. This reduction is mainly attributed to removing duplicate districts per year, as discussed in the previous paragraph. The STATA syntax utilised is detailed in Appendix F for reference.

Appendix A contains a correlation matrix encompassing the variables under examination in the analysis. Correlation, as investigated by Brooks and Tsolacos (2010), is measured on a scale ranging from -1 to 0 or from 0 to 1. A value of -1 denotes a robust negative correlation between variables, while a value of +1 indicates a pronounced positive correlation between variables. This analytical approach facilitates a nuanced understanding of the interrelationships between the variables under examination. In this case, it can be seen that there is a correlation between vacancy levels per 1,000 inhabitants and most of the independent variables. From all the age groups, the strongest correlation can be found between the dependent variable and the age group 65+. There is also a correlation between the dependent variable and the control variable for urbanity level (0.177).

The table below shows the descriptive statistics for all variables included.

Table 2: Descriptive statistics for all variables

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Vacancy level per 1,000 inh.	15994	.578	1.397	0	13.721
Number of inhabitants	15994	5453.668	7109.603	5	109805
Share of 0-14 years	15994	3.371	1.904	0	34.337
Share of 15-24 years	15994	2.727	1.949	0	100
Share of 25-44 years	15994	5.153	3.915	0	100
Share 45-64 years	15994	6.616	3.167	0	100
Share of 65+ years	15994	4.587	2.528	0	100
Average standardised income	9759	34.55	7.061	6.3	150.7
Mean WVO	7793	222.116	328.51	10	7134
Mean year of construction	6226	1957.79	38.141	1625	2021
Dummy for urbanity level	15994	.388	.487	0	1

Table 2 highlights the differences in observations. For instance, the number of observations for income, WVO (average residential area), and average construction year differ. The discrepancy in income data is because the year 2023 contains only missing values, as this data was not available. The missing values in construction year data are due to the absence of data for 2019, as previously discussed in the prior paragraph. The lower number of observations for the average WVO and also the average construction year is because this data comes from Locatus and was only available for vacant properties and thus districts with vacant buildings. The other data, from CBS, includes 15,994 observations covering all districts per year, including those districts without vacant properties. Replacing these missing values with “0” was not an option as it would skew the averages and affect the regression analyses. Additionally, the minimum value of the dependent variable is 0. To examine the number of zeroes in the dependent variable, a table is provided in Appendix A. The histogram of the dependent variable in Appendix C also shows the frequency of zeroes. It shows that 8,206 districts have zero vacant buildings per 1,000 inhabitants. Furthermore, Table 5 shows that there are on average 0.578 vacant retail premises per 1,000 inhabitants, with a standard deviation of 1.397, indicating variability across districts. The maximum value of 13.721 suggests some areas have significantly higher vacancy rates.

Furthermore, by generating histograms of all variables, the skewness was checked. Some variables, such as the number of inhabitants and the average WVO, were not normally distributed. Logarithms were applied to these variables to reduce skewness, resulting in a better approximation of a normal distribution (Brooks & Tsolacos, 2010). Additionally, a box plot was created for the dependent variable, revealing the presence of outliers. Consequently, the top 1% of values were removed to minimise the impact of these outliers on the analyses.

To provide an illustrative representation of the levels of vacancy per district in the Netherlands and to visualise changes over the years, GIS maps have been created. Figure 9 en 10 illustrate the vacancy levels for the years 2019 and 2023. Additional maps for the years 2020, 2021, and 2022 are included in Appendix B. Here, only the beginning (2019) and the end (2023) of a “cycle” as discussed in Chapter 1 are presented to highlight the differences.

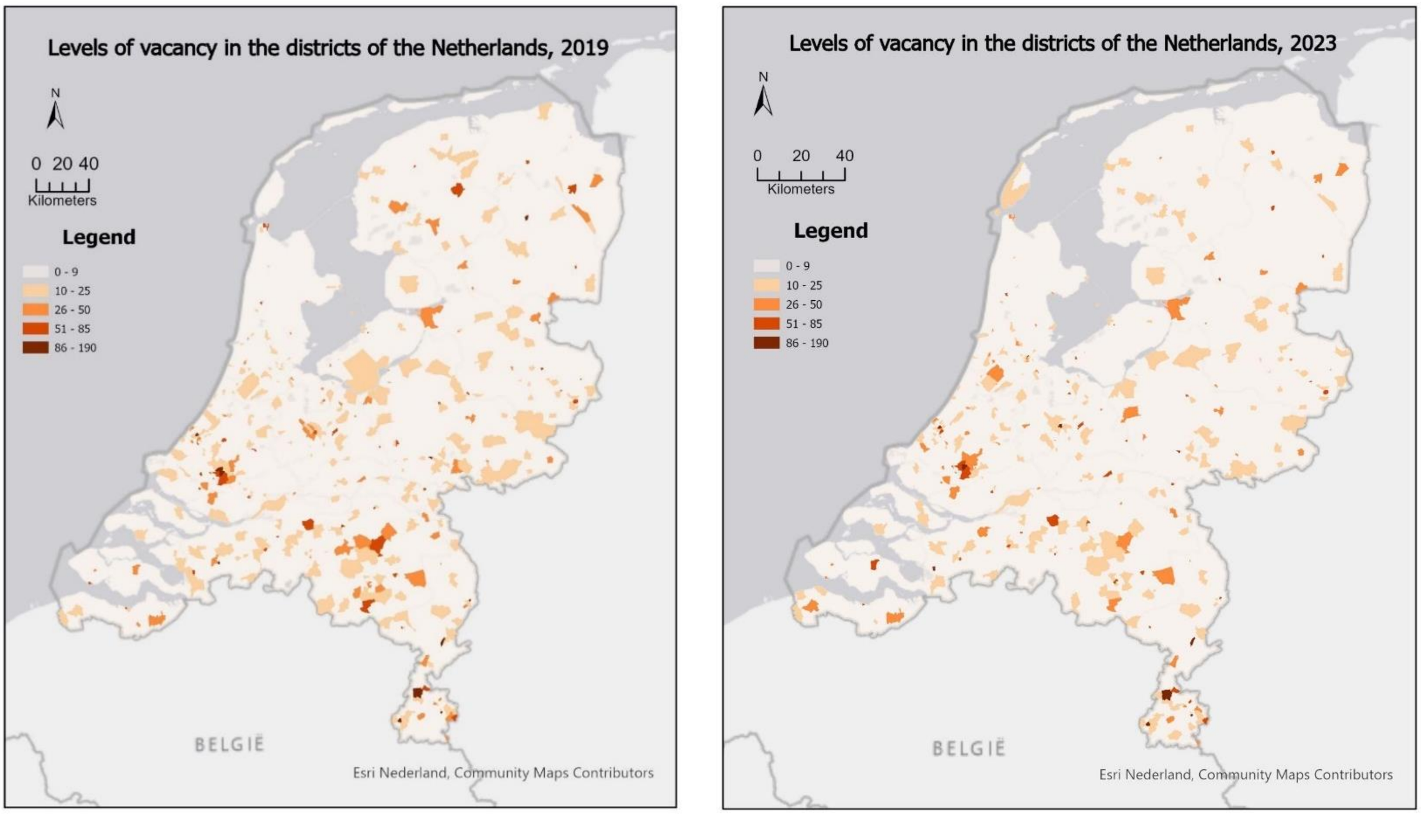

Figure 10: Map of levels of vacancy in the districts of the Netherlands, 2019 and 2023

Comparing the five maps, the vacancy has visually decreased. Vacancy has diminished notably in Flevoland, while a slight increase is observed near Texel. Additionally, there is also a noticeable reduction in vacancies in Friesland and Groningen. The significant differences between provinces must be considered in the empirical models presented in Chapter 4.

3.4 Methodology

The dependent variable, denoted as y, represents the continuous measure of the number of vacant retail premises per 1,000 inhabitants at the district level. Similarly, the independent variables are continuous and also evaluated at the district level. Given the continuous nature of the dependent variable, linear regression is deemed suitable for analysis. Multiple regression models focus on retail vacancy levels per 1,000 inhabitants and district characteristics for all years from 2019 to 2023, with variables being gradually added.

In this study, multiple independent variables are employed to investigate the variance in the dependent variable y. Consequently, a multiple linear regression model is employed to quantify the impact of these independent variables on y. In this model, y serves as the response variable, with $\beta 0$ representing the intercept, $\beta 1$ denoting the slope, x signifying the predictor or regressor variable, and ϵ indicating the error term, which encapsulates the disparity between the observed value of y and the linear relationship ($\beta 0 + \beta 1x$) (Montgomery et al., 2012). The multiple linear regression model yields the following regression equation, predicated on the variables under analysis.

$$RVjt = \beta 0 + \beta 1NCjt + \beta 2A1jt + \beta 3A2jt + \beta 4A3jt + \beta 5A4jt + \beta 6A5jt + \beta 7HIjt + \beta 8PSjt + \beta 9LCjt + \beta 10BYjt + \gamma t + \mu j + \varepsilon jt \quad (1)$$

The dependent variable RV represents the number of vacant retail premises per 1,000 inhabitants in each district j during year t. NC signifies the population count within area j during year t. A1 corresponds to the population count in the age group 0-15 years within district j in year t. A2 signifies the population count in the age group 15-25 years within district j in year t. A3 denotes the population count in the age group 25-45 years within district j in year t. A4 represents the population count in the age group 45-65 years within district j in year t. A5 indicates the population count in the age group 65 or older within district j in year t. HI represents the mean standardised income in district j during year t, measured in thousands of euros. PS denotes the property size in district j during year t, measured in WVO. LC is a dummy variable indicating the urbanity level of each district j during year t, distinguishing between rural and urban areas. BY denotes the year of construction for each district j during year t. γt controls for time fixed effects and represents the different time periods in year t. μj are dummy variables for the twelve provinces in the Netherlands, controlling for the spatial fixed effects in area j. ϵjt is the error term.

The linear regression analysis rests upon several critical assumptions, each of which warrants examination to uphold the integrity of the model. Firstly, it is essential to ensure the independence of the independent variables, indicating the absence of multicollinearity issues. This can be evaluated through the Variance Inflation Factor (VIF), with a VIF score surpassing five indicative of potential multicollinearity concerns. The VIF scores can be found in Appendix C, where it is evident that no multicollinearity issues arise. All VIF scores are below 5, except for those of certain province dummies. This is not problematic and occurs because the inclusion of spatial fixed effects (province dummies) inherently introduces multicollinearity. These dummies capture the unique effects of each province, which may be highly correlated with other variables or with each other.

Secondly, it is imperative to verify the linearity between the dependent variable and the independent variables. This can be achieved by constructing two-way scatterplots and fitting linear prediction lines to assess the linearity of the relationship. The graphs in Appendix C show linear relationships so this assumption is fulfilled.

The third assumption revolves around the normality of residuals, positing that residuals should ideally follow a random or normal distribution. The Shapiro-Wilk test can be used to assess this assumption, with a p-value below 5% indicative of deviation from normality. The outcomes in Appendix C show that the data is not normally distributed. But because of the size of the research, the test results will still be reliable, because the normal approximation to the sample distribution of W is valid for sample sizes between 4 and 2000.

Moreover, adherence to the assumption of constant error variance, or homoscedasticity, is crucial. This entails ensuring that the variance of residual terms remains consistent across all levels of the predictor variables. Any departure from this assumption, known as heteroskedasticity, can be evaluated using the Breusch-Pagan heteroskedasticity test, where a p-value below the significance level suggests the presence of heteroskedasticity (Brooks & Tsolacos, 2010). The outcome of this test in Appendix C shows that heteroskedasticity is detected here. However, it can be mitigated by estimating the regression model with robust standard errors.

4. RESULTS

4.1 Findings OLS model

Table 3 presents the results for Models 1, 2, 3, 4 and 5. The first model includes the independent variables: the number of inhabitants and the age groups. The formula for this model is presented below.

$$RVjt = \beta 0 + \beta 1NCjt + \beta 2A1jt + \beta 3A2jt + \beta 4A3jt + \beta 5A4jt + \beta 6A5jt + \varepsilon jt \qquad (2)$$

A logarithm of the number of inhabitants has been taken, and shares have been used for the age groups (see Chapter 3). This model has 15,994 observations. The R-squared for this model is 0.0629, meaning that approximately 6.29% of the variance in the dependent variable (vacancy per 1,000 inhabitants) is explained by the independent variables in the model, namely the number of inhabitants and the share of age groups. This suggests that the model has a low explanatory power, indicating that most of the variability is influenced by factors not included in the model. Therefore, additional variables are incorporated in Models 2, 3, 4, and 5 to improve the explanatory power.

In Model 1, a 1% increase in the number of inhabitants leads to an increase of 0.106 in retail vacancy per 1,000 inhabitants, with a significance level of 99%. This finding is not in line with Hollander et al. (2018). However, regional differences have not yet been accounted for in this model, so this represents the effect across the entire country. In Models 4 and 5, where regional differences are included, there is a negative association. This suggests that the positive association observed in the national trend does not reflect regional variations, such as the structure of a region and the relationship between basic and non-basic sectors, as discussed in Chapter 2 (Wang and Hofe, 2008). A province with a diverse economy might absorb population increases more effectively without leading to higher retail vacancy rates.

Furthermore, among the age groups, a 1 percentage point increase in the share of 0-14 year-olds leads to a decrease of 0.191 in retail vacancy per 1,000 inhabitants, with a significance level of 99%. Conversely, a 1 percentage point increase in the share of those aged 65+ leads to an increase of 0.057 in retail vacancy per 1,000 inhabitants, also with 99% significance. A possible explanation could be that people aged 65+ consume less than younger households in the Netherlands (CBS, 2017). However, it contradicts the finding that older households spend relatively larger portions of their earnings compared to younger age groups. A possible explanation is that they may be spending more online or on other things rather than retail purchases. Additionally, the share of the 15-24 age group in this model is not significant at the 95% level, as $p > 0.05$.

Model 2 controls for time, resulting in the same number of observations as Model 1 and a slightly higher R-squared (0.0637). The corresponding equation for this model is presented below.

$$RVjt = \beta 0 + \beta 1NCjt + \beta 2A1jt + \beta 3A2jt + \beta 4A3jt + \beta 5A4jt + \beta 6A5jt + \gamma t + \varepsilon jt \qquad (3)$$

Additionally, the coefficients for population size and age groups remain largely consistent compared to Model 1. This indicates that the association between population growth and age groups on the dependent variable is not significantly affected by the specific year. All year dummies are not significant except for

2023, which shows a significant decrease (with 99% confidence) in retail vacancy per 1,000 inhabitants compared to the reference year (2019). This is also confirmed by the GIS maps of 2019 and 2023 presented in Chapter 3, which show a clear reduction in vacancy levels.

In the equation for Model 3, it can be seen that a control variable for urbanity level is added.

$$RVjt = \beta 0 + \beta 1NCjt + \beta 2A1jt + \beta 3A2jt + \beta 4A3jt + \beta 5A4jt + \beta 6A5jt + \beta 9LCjt + \gamma t + \varepsilon jt \quad (4)$$

The number of observations remains the same, and the R-squared increases slightly to 7.7%, indicating that now 7.7% of the variance is explained by the independent variables. The coefficients for the age groups remain relatively consistent compared to Models 1 and 2. However, the coefficient for the number of inhabitants drops from approximately 0.105 in the first two models to about 0.045 in this third model, indicating that the relationship between population size and retail vacancy is weaker when urbanisation is included in the model. This suggests that population growth has less impact on retail vacancy in urban areas than in rural areas. This aligns with Christaller's theory, as urban areas (central places) have a higher concentration of amenities and economic activities, making them better able to absorb population growth (Peek & Veghel, 2011). Additionally, the urbanisation dummy is positively associated (0.400) with retail vacancy per 1,000 inhabitants, which is consistent with the findings of Remøy et al. (2007) and Oskam (2021) that vacancy rates are higher in more urbanised areas than in rural areas. Possible explanations for this include higher rental prices and greater competition in urban areas.

Table 3: Baseline regression results

Variables	1	2	3	4	5
Ln (number of citizens)	0.106***	0.105***	0.045***	-0.191***	-0.214***
	(0.010)	(0.010)	(0.011)	(0.013)	(0.020)
Age groups (in %)[1]					
0-14 years	-0.191***	-0.190***	-0.182***	-0.124***	-0.298***
	(0.014)	(0.014)	(0.013)	(0.010)	(0.022)
15-24 years	0.026*	0.026*	0.024*	0.024**	0.077***
	(0.014)	(0.014)	(0.013)	(0.012)	(0.027)
25-44 years	0.052***	0.053***	0.039***	0.032***	0.120***
	(0.008)	(0.008)	(0.007)	(0.007)	(0.016)
45-64 years	-0.034***	-0.035***	-0.027***	-0.017**	-0.043***
	(0.010)	(0.010)	(0.009)	(0.007)	(0.014)
65+ years	0.057***	0.059***	0.061***	0.042***	0.088***
	(0.014)	(0.014)	(0.014)	(0.012)	(0.015)
Year dummies					
(Reference category: 2019)					
2020		0.007	0.008	0.013	-0.002
		(0.034)	(0.034)	(0.031)	(0.030)
2021		-0.013	-0.012	-0.010	-0.033
		(0.034)	(0.034)	(0.031)	(0.031)
2022		-0.061*	-0.063*	-0.049	-0.112**
		(0.033)	(0.033)	(0.030)	(0.049)
2023		-0.095***	-0.088***	-0.096***	
		(0.033)	(0.033)	(0.030)	

[1] Note that a 1 percentage point increase in the proportion of elderly households necessarily implies a corresponding 1 percentage point decrease in the proportion of other age groups combined. This interdependence must be taken into account when interpreting the results.

Urbanity level			0.400***	0.439***	0.353***
(Reference category: rural)			(0.030)	(0.029)	(0.034)
Location					
(Reference category: Flevoland)					
Drenthe				1.417***	1.307***
				(0.125)	(0.133)
Friesland				1.167***	1.168***
				(0.054)	(0.069)
Gelderland				1.196***	1.149***
				(0.058)	(0.066)
Groningen				1.148***	1.089***
				(0.078)	(0.097)
Limburg				1.416***	1.380***
				(0.053)	(0.064)
Noord-Brabant				1.499***	1.280***
				(0.063)	(0.059)
Noord-Holland				1.153***	1.155***
				(0.060)	(0.071)
Overijssel				1.383***	1.301***
				(0.074)	(0.079)
Utrecht				1.419***	1.295***
				(0.104)	(0.105)
Zeeland				1.754***	1.696***
				(0.096)	(0.118)
Zuid-Holland				1.162***	1.104***
				(0.049)	(0.053)
Income					
Mean standardised income					-0.001
					(0.001)
Constant	0.015	0.044	0.351***	1.402***	1.677***
	(0.084)	(0.088)	(0.091)	(0.098)	(0.161)
Observations	15,994	15,994	15,994	15,994	9,759
R-squared	0.063	0.064	0.077	0.240	0.261

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Model 4 controls for location by adding province dummies to the model. The corresponding formula is presented below.

$$RVjt = \beta 0 + \beta 1NCjt + \beta 2A1jt + \beta 3A2jt + \beta 4A3jt + \beta 5A4jt + \beta 6A5jt + \beta 9LCjt + \gamma t + \mu j + \varepsilon jt \qquad (5)$$

The number of observations remains the same, but the R-squared is significantly higher: 24.03% of the variance is now explained by the independent variables, which is about four times as much as in the first model. Notably, the number of inhabitants suddenly negatively associates with retail vacancy per 1,000 inhabitants. This means that when considering the effect within provinces instead of across the entire Netherlands, an increase in population actually leads to a decrease in retail vacancy per 1,000 inhabitants. Compared to Model 3, the negative association for the 0-14 age group and the positive associations for the 45-64 and 65+ age groups are slightly less pronounced.

In the equation below it can be seen that Model 5 adds household income (average standardised income) to Model 4.

$$RVjt = \beta 0 + \beta 1NCjt + \beta 2A1jt + \beta 3A2jt + \beta 4A3jt + \beta 5A4jt + \beta 6A5jt + \beta 7HIjt + \beta 9LCjt + \gamma t + \mu j + \varepsilon jt \qquad (6)$$

The first noticeable change is the drop in observations from 15,994 to 9,759. This is because information on average standardised income is unavailable for 2023 (see Chapter 3), so 2023 is excluded from this model. Additionally, the coefficient for income is not significant, indicating that this variable has no significant impact on vacancy rates. This contradicts the findings of Bieniek et al. (2018) and Kickert (2021), who assert that changes in income distribution have a noticeable influence on retail vacancy.

In Appendix D, more relevant variables are added to check the robustness of the results. The significance of the coefficient for average standardised income therefore changes in Model 6, where the average WVO is included in the regression. It is notable that an increase of 1000 euros in income results in a rise of 0.0264 in retail vacancy per 1,000 inhabitants. This contradicts Kang (2019), who found an association between higher incomes and higher retail turnover. Adding the control variable for retail property size leads to the conclusion that a 1% increase in size results in a 0.2453 increase in vacancy per 1,000 inhabitants. This indicates that larger retail properties are associated with higher vacancy rates, which contradicts Ball (2002) from the theoretical framework, who suggested that smaller buildings are more frequently vacant. Research by Custers (2022) indicates that larger properties do tend to have higher vacancy rates, but the largest properties have relatively lower vacancy rates. This model, however, contains 5,456 observations and is included in the Appendix due to many missing values. This also applies to Model 7, where the control variable for building year is added. As a result, the observations drop to 3,930, but the R-squared increases to 30.7%. The observations are so low because the average building year was not available for 2019 (see Chapter 3). This means that this regression only pertains to 2020, 2021, and 2022. When the average building year increases by one year, the vacancy per 1,000 inhabitants decreases by 0.0032. This indicates that newer buildings are associated with lower vacancy rates. This aligns with the findings of Yakubu et al. (2017), who stated that older buildings have higher vacancy rates. A possible explanation for this is the sustainability ambitions for buildings in the Netherlands (Ministerie van Algemene Zaken, 2024). Older buildings are often less well-insulated and consequently have lower energy ratings. Since a minimum energy rating of C is already required for offices, it is very likely that this requirement will eventually apply to retail properties as well. Additionally, a higher energy rating often equates to lower energy costs.

Based on all models, the hypotheses framework from Chapter 2 can be revised. Regarding age composition, there are both positive and negative associations because the results do not offer unequivocal support for this hypothesis.

Table 4: Hypotheses of the independent variables, revised

Independent Variable	*Expected sign*	*Actual sign*
Age composition	-	- v +
Population size	-	-
Household income	-	+

4.2 Sensitivity analysis

Urbanity level

In Chapter 2, it was anticipated that location (urban or rural) would have a positive association with retail vacancy per 1,000 inhabitants, meaning that there would be higher vacancy rates in urban areas compared to rural areas. This expectation is confirmed in section 4.1. To determine whether the effects of the explanatory variables—population size and age groups—differ across levels of urbanization, a Chow F-test was conducted. The independent variables from model 4 were included in this analysis. This test compares the observations of two groups, the "unrestricted models", and examines if there are differences in regression results between these groups. The test was conducted based on the restriction "urbanisation_dummy", which can be either 0 or 1. In this study areas with up to 1000 addresses per km2 are categorised as rural (0), and those with more than 1000 addresses per km2 are categorised as urban (1). When examining the results of these separate regressions, it is notable that the association between retail vacancy per 1,000 inhabitants and population size is lower in urban areas compared to rural areas. This means that in urban areas, an increase in population size leads to a greater decrease in vacancy compared to rural areas. This finding is inconsistent with Christaller's theory, as previously discussed in section 4.1. The associations between all age groups and retail vacancy also show a stronger influence in urban areas compared to rural areas. Notably, only the coefficient for the youngest age group, 0-14 years, is significant in rural areas. In urban areas, a one percentage point increase in the share of the 65+ population leads to a 0.123 increase in vacancy per 1,000 inhabitants. This indicates a greater impact of the elderly population on retail vacancy in cities compared to rural areas. After running the unrestricted models, a pooled restricted model was conducted. Both unrestricted and restricted models are presented below in table 5.

Table 5: Regression results unrestricted and restricted models

	(1)	(2)	(3)
Variables	Pooled	Rural	Urban
Ln (number of citizens)	-0.128***	-0.150***	-0.238***
	(0.008)	(0.007)	(0.021)
Age groups (in %)			
0-14 years	-0.134***	-0.026***	-0.209***
	(0.008)	(0.008)	(0.016)
15-24 years	0.028***	0.005	0.043***
	(0.006)	(0.006)	(0.013)
25-44 years	0.047***	0.008	0.044***
	(0.003)	(0.005)	(0.006)
45-64 years	-0.025***	-0.003	-0.045***
	(0.005)	(0.004)	(0.014)
65+ years	0.041***	-0.006	0.123***
	(0.005)	(0.004)	(0.011)
Year dummies			
(Reference category: 2019)			
2020	0.012	0.031	0.002
	(0.031)	(0.026)	(0.067)
2021	-0.011	-0.016	0.024
	(0.031)	(0.026)	(0.067)
2022	-0.047	-0.037	-0.037
	(0.031)	(0.026)	(0.066)
2023	-0.104***	-0.065**	-0.118*
	(0.031)	(0.026)	(0.067)

Location			
(Reference category: Flevoland)			
Drenthe	1.308***	1.131***	2.472***
	(0.078)	(0.057)	(0.238)
Friesland	1.074***	1.113***	1.148***
	(0.067)	(0.051)	(0.180)
Gelderland	1.157***	0.967***	1.414***
	(0.045)	(0.040)	(0.091)
Groningen	1.067***	1.065***	1.180***
	(0.075)	(0.060)	(0.177)
Limburg	1.391***	1.205***	1.702***
	(0.048)	(0.041)	(0.102)
Noord-Brabant	1.489***	1.488***	1.480***
	(0.038)	(0.034)	(0.077)
Noord-Holland	1.223***	1.086***	1.214***
	(0.045)	(0.055)	(0.074)
Overijssel	1.354***	1.281***	1.455***
	(0.053)	(0.046)	(0.108)
Utrecht	1.483***	1.348***	1.512***
	(0.059)	(0.069)	(0.097)
Zeeland	1.662***	1.715***	1.786***
	(0.068)	(0.050)	(0.212)
Zuid-Holland	1.249***	1.017***	1.270***
	(0.039)	(0.048)	(0.065)
Constant	1.085***	1.104***	2.171***
	(0.067)	(0.057)	(0.182)
Residual Sum of Squares	24,165.920	6,373.163	16,607.448
Observations	15,994	9,785	6,209
R-squared	0.225	0.312	0.208

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The formula and calculation of the Chow F-test is detailed in Appendix E. RSSp, RSS1, and RSS2 represent the residual sum of squares for each model (in this case, RSSp = 24165.920, RSS1 = 6373.163, and RSS2 = 16607.448). Here, k is the number of parameters, which is 21 plus an intercept. Additionally, N1 and N2 are the observations in each group (N1 = 9785 and N2 = 6209). The test statistic is $F(22, 15950) = 37.39$ and the 5% critical value is $F[22, \infty] = 1.543$. The null hypothesis (H0) that the coefficients are the same for the two models can be rejected with 95% confidence. This indicates a significant structural change in the regression parameters between rural and urban areas.

An additional test with interaction terms was conducted in STATA. In this restricted pooled model, multiple interaction terms were included to test whether the relationships between the explanatory variables and the dependent variable differ based on urbanisation. This model with the associated interaction terms are detailed in Appendix E. The specific hypothesis tested here is whether the interaction effects of urbanisation with the various age share variables and population count are equal to zero. The test results show an F-statistic of 66.79 with a p-value of 0.0000. This means that the null hypothesis (H0), that all coefficients of the interaction terms are equal to zero, is rejected with 99% confidence. Therefore, the effects of the independent variables on the dependent variable differ based on urbanisation. Urbanisation thus plays a significant role in how population size and age groups influence retail vacancy per 1,000 inhabitants. Hence, H2 from Chapter 2, which posits that demographic factors will contribute to retail vacancy disparities between rural and urban districts in the Netherlands, cannot be rejected based on the results.

COVID-19

Given that this study examines the period from 2019 to 2023, it is interesting to investigate whether the year affects the relationship between the independent variables and the dependent variable. This can strengthen the answer to the main research question. To this end, another Chow F-test is performed. Because this analysis does not involve two distinct groups, as in the case of urbanisation (0 = rural and 1 = urban), the years are divided into three groups to maintain the robustness of the conclusion. The groups are divided as follows: 2019, 2020 + 2021, and 2022 + 2023. The first group represents the year before the onset of COVID-19, the second group represents the peak of the pandemic, and the last group represents the aftermath (Ministerie van Algemene Zaken, 2024a). Both unrestricted and restricted models are presented in table 6.

Table 6: Regression results unrestricted and restricted models

	(1)	(2)	(3)	(4)
Variables	Pooled	Before	During	After
Ln (number of citizens)	-0.127***	-0.137***	-0.156***	-0.116***
	(0.008)	(0.018)	(0.014)	(0.012)
Age groups (in %)				
0-14 years	-0.135***	-0.127***	-0.267***	-0.098***
	(0.008)	(0.017)	(0.017)	(0.010)
15-24 years	0.027***	0.020*	0.047**	0.027***
	(0.006)	(0.010)	(0.019)	(0.009)
25-44 years	0.046***	0.066***	0.126***	0.029***
	(0.003)	(0.009)	(0.011)	(0.004)
45-64 years	-0.024***	-0.023**	-0.055***	-0.022***
	(0.005)	(0.010)	(0.012)	(0.006)
65+ years	0.039***	0.040***	0.070***	0.034***
	(0.005)	(0.011)	(0.011)	(0.006)
Location (Reference category: Flevoland)				
Drenthe	1.311***	1.154***	1.374***	1.370***
	(0.078)	(0.162)	(0.124)	(0.125)
Friesland	1.074***	1.387***	1.079***	0.935***
	(0.067)	(0.143)	(0.109)	(0.106)
Gelderland	1.157***	1.219***	1.205***	1.102***
	(0.045)	(0.098)	(0.073)	(0.072)
Groningen	1.069***	1.218***	1.106***	0.949***
	(0.075)	(0.161)	(0.117)	(0.121)
Limburg	1.391***	1.374***	1.439***	1.338***
	(0.048)	(0.103)	(0.079)	(0.075)
Noord-Brabant	1.488***	1.485***	1.546***	1.419***
	(0.038)	(0.085)	(0.063)	(0.059)
Noord-Holland	1.221***	1.214***	1.345***	1.177***
	(0.045)	(0.099)	(0.076)	(0.068)
Overijssel	1.354***	1.486***	1.385***	1.284***
	(0.053)	(0.113)	(0.086)	(0.083)
Utrecht	1.485***	1.409***	1.656***	1.352***
	(0.059)	(0.123)	(0.094)	(0.096)
Zeeland	1.660***	1.840***	1.746***	1.519***
	(0.068)	(0.160)	(0.110)	(0.102)
Zuid-Holland	1.248***	1.334***	1.327***	1.127***
	(0.039)	(0.085)	(0.064)	(0.060)

Constant	1.060***	1.028***	1.337***	0.946***
	(0.064)	(0.138)	(0.121)	(0.095)
Residual Sum of Squares	24,193.174	4,281.823	9,993.060	9,619.585
Observations	15,994	3,103	6,317	6,574
R-squared	0.224	0.243	0.253	0.208

Standard errors in parentheses
*** p<0.01, ** p<0.05, *p<0.1

When comparing the coefficients of the three groups, it is evident that the negative association between population size and retail vacancy per 1,000 inhabitants is fairly consistent. However, this is not the case for the age groups. It is notable that the age share of 0-14 years has a strong negative effect on vacancy rates, with the effect being significantly greater in 2020-2021 (-0.267) and significantly smaller in 2022-2023 (-0.098) compared to 2019 (-0.127). The difference between the peak period of COVID-19 and its aftermath is noticeable. The stronger influence of the younger population on retail vacancy during COVID-19 may be explained by the fact that this group did not shop online and continued to purchase items in physical stores. The age group of 15-24 years shows a positive association with vacancy rates, with the association being less significant in 2019 than in the years thereafter. This implies a changing influence of young adults during and after COVID-19. Additionally, the 25-44 age group has a positive association with vacancy rates in 2019 (0.066) and even stronger in 2020-2021 (0.126), while this association weakens in 2022-2023 to 0.029. This may indicate that this group shopped online extensively before and during COVID-19 and later had a renewed interest in physical stores. Furthermore, it is noteworthy that the 45-64 age group has a negative association with vacancy rates, which is strongest during the COVID-19 period, while the 65+ age group has a positive association with vacancy rates, also strongest during that period. This is because, due to COVID-19, stores were substantially less visited, especially by more vulnerable groups (Ministerie van Infrastructuur en Waterstaat, 2022). This may indicate that elderly people visit stores less frequently than younger people. As the demand for physical stores decreased, and despite substantial government assistance to store owners, the results show that retail vacancy rates especially increased in locations with ageing populations.

The calculation of the Chow F-test is detailed in Appendix E. RSS_p, RSS_1, and RSS_2 represent the residual sum of squares for each model (in this case, $RSS_p = 24193.174$, $RSS_1 = 4281.823$, $RSS_2 = 9993.060$ and $RSS_3 = 9619.585$). Here, k is the number of parameters, which is 17 plus an intercept. Additionally, N_1, N_2 and N_3 are the observations in each group ($N_1 = 3103$, $N_2 = 6317$ and $N_3 = 6574$). The test statistic is $F(18, 15940) = 11.072$ and the 5% critical value is $F[18, \infty] = 1.57$. The null hypothesis (H0) that the coefficients are the same for the two models can be rejected with 95% confidence. This indicates a significant structural change in the regression parameters between the three time groups.

Subsequently, a pooled regression model with interaction terms was conducted in STATA. The results are also included in Appendix E. The F-value is 13.39, indicating a significant difference in the regression coefficients between the different groups, and the null hypothesis can be rejected with 99% confidence (p-value = 0.000). These results suggest that the relationship between population size, age groups, and retail vacancy per 1,000 inhabitants varies over different time periods. Based on these results, H3, which posits that demographic factors contribute to differences in retail vacancy before, during, and after COVID-19, cannot be rejected.

5. CONCLUSION

5.1 Conclusion

This thesis aims to answer the research question: "To what extent are demographic factors associated with retail vacancy in the districts of the Netherlands?" An answer to this question can be derived from the results of this research. The motivation for this study was formed by the increasing number of individuals aged 65 and over, ABN AMRO's prediction that this would lead to a 2% loss in revenue, and the trend of declining vacancy rates (ABN AMRO, 2022; Custers, 2023). Literature review revealed that older adults have less disposable income but spend more of it, and that the social aspect of shopping is very important to this group (Brenner and Clarke, 2017; CBS, 2017; Lesáková, 2016). This led to the hypothesis that a higher age group would negatively associate with vacancy per 1,000 inhabitants. Similarly, for the other examined demographic factors, population growth and household income, a negative association with vacancy was derived from theory.

The hypothesis regarding population growth turns out to be the only one confirmed. This aligns with the Retail Gravitation Theory, which states that the population size in a region influences the attractiveness of a shopping area (Reilly, 1931; Brown, 1992). It may also suggest that there is a strong basic sector at the provincial level that increases the demand for shops, as the Economic Base Theory posits (Wang and Hofe, 2008). The hypothesis that income negatively associates with retail vacancy per 1,000 inhabitants seems incorrect; however, this result was not significant. This suggests that there is no relationship, but it can be seen that this relationship does become significant when control variables such as average building year and retail floor space (WVO) are added. Here, a positive association is notable, which contradicts the theories from chapter 2, including those by Kang (2019), who concluded that there is an association between higher incomes and higher retail turnover.

The hypothesis regarding different age groups cannot be definitively confirmed or refuted. The youngest age group (0-14 years) associates negatively with vacancy per 1,000 inhabitants, while the oldest age group (65+) associates positively with vacancy per 1,000 inhabitants. The only other group that negatively associates with vacancy is the 45-64 age group. Thus, it cannot be concluded that the older the age, the more negative the association with vacancy. However, it can be stated that locations with an ageing population have higher retail vacancy per 1,000 inhabitants than locations with a high proportion of young population (0-14 years). To address the question posed by Schröder (2023) mentioned in the introduction: ageing appears to be more of a threat to the retail sector than an opportunity.

In examining these demographic factors and their influence on retail vacancy, considering urban and rural differences was essential. The findings of Remøy et al. (2007) and Oskam (2021) that retail vacancy is higher in urban areas than in rural areas can be confirmed. Furthermore, the association between demographic factors and retail vacancy per 1,000 inhabitants also differs between urban and rural areas. This highlights the importance of considering this geographic aspect in analysing to what extent demographic factors associate with retail vacancy.

In addition to considering geographic differences, the impact of the COVID-19 pandemic cannot be ignored due to the timeframe studied. The most recent vacancy cycle, which started in 2019, already showed in

Figure 1 a slight increase in vacancy during the COVID-19 years 2020 and 2021, after which vacancy slightly decreased again in 2022 (Slob, 2023). This research also shows that the association between demographic factors and retail vacancy per 1,000 inhabitants differs in the periods before, during, and after COVID-19. This underscores the importance of including this aspect in the analysis and calls for further research into the precise impact of this pandemic on retail vacancy.

5.2 Advice for real estate practitioners

Based on the conclusions of this research, it is important for professionals in this field to anticipate these findings. For policymakers, the conclusion that an ageing population could potentially lead to increased vacancy is crucial for shaping policies for shopping areas, considering the growing elderly population in the Netherlands. This could lead to policies that encourage older adults to shop more or focus on stores in areas with a high proportion of the 0-14 age group. Additionally, policies could focus on expanding shopping areas with high population densities, as these areas experience less vacancy. Furthermore, it is important to consider rural and urban differences, which could manifest in preserving stores in rural areas and potentially redeveloping vacant properties in urban areas.

Investors, managers, and other practitioners in this real estate sector should also respond to these findings. Investors might find it more advantageous to invest in shopping areas with growing populations and regions with a large proportion of the youngest age group, while being cautious with investments in areas with an aging population. Additionally, it is beneficial to consider the size of retail spaces and ensure they meet user needs, as the results indicated that WVO positively associated with retail vacancy per 1,000 inhabitants. This also applies to the building year, where focusing on newer buildings might be wiser in terms of sustainability and energy costs (Ministerie van Algemene Zaken, 2024). For managers, considering more flexible leasing terms in the aftermath of COVID-19 could be beneficial.

By implementing the conclusions from this research, real estate practitioners can better anticipate changes affecting retail vacancy, allowing them to address potential problems early and seize interesting opportunities as they arise.

5.3 Limitations and further research

The limitations of this study primarily lie in the number of observations and the missing data for certain variables, such as average standardised incomes for the year 2023 and the average building year for the year 2019. Another limitation is the timeframe, which spans only five years, causing the three time period groups for the Chow F-test for COVID-19 to contain unequal numbers of years, as 2019 is the only “pre-COVID” year. Additionally, the Locatus data only includes information on vacant retail properties, not on properties that are not vacant. As a result, it was not possible to calculate a vacancy rate for the districts, only a level of vacancy per 1,000 inhabitants. This also limited the number of observations, as adding variables like average building year and WVO immediately excluded all districts without vacant properties. Two final limitations of this study are that it does not distinguish between types of vacancy as defined by Evers et al. (2014) and Van Gool et al. (2007), and it does not account for changes in the level of retail vacancy within a year, as Locatus measures it only once annually.

Based on the conclusions drawn, the advice for real estate practitioners, and the limitations, several suggestions for future research arise. Future studies could delve deeper into the impact of COVID-19 on retail vacancy or investigate the causes of rural and urban differences in retail vacancy. Additionally, it would be interesting to examine the influence of household income on retail vacancy once the 2023 income data becomes available from CBS. This could involve analysing the data over the past five years and including other variables to determine if a significant association emerges. Future research could also consider the differences in types of vacancy when conducting the same analysis as in this study.

LITERATURE LIST

ABN AMRO (2022) 'De oudere consument van morgen,' *ABN AMRO*, 6 September. https://www.abnamro.nl/nl/zakelijk/insights/sectoren-en-trends/retail/de-oudere-consument-van-morgen.html.

Adhvaryu, B. (2010) 'Enhancing urban planning using simplified models: SIMPLAN for Ahmedabad, India', *Progress in Planning*, 73(1), pp. 113-207.

Alonso, W. (1964) *Location and land use: Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

Ball, R.M. (2002) 'Re-use potential and vacant industrial premises: revisiting the regeneration issue in Stoke-on-Trent', *Journal of Property Research*, 19(2), pp. 93-110.

Benjamin, J.D., Jud, G.D. and Winkler, D.T. (1998) 'A simultaneous model and empirical test of the demand and supply of retail space,' *Journal of Real Estate Research*, 16(1), pp. 1–14. https://doi.org/10.1080/10835547.1998.12090938.

Bieniek, C., Lobaugh, K., Pincha, P. and Stephens, B. (2018) *The Great Retail Bifurcation – Why the Retail 'Apocalypse' Is Really a Renaissance*. Deloitte Insights Series. New York: Deloitte.

Bolt, E.J. (2003) *Winkelvoorzieningen op waarde geschat; theorie en praktijk*. Merkelbeek: Drukkerij Bakker.

Brenner, A.B. and Clarke, P. (2017) 'Difficulty and independence in shopping among older Americans: more than just leaving the house,' *Disability and Rehabilitation*, 41(2), pp. 191–200. https://doi.org/10.1080/09638288.2017.1398785.

Brooks, C. & Tsolacos, S. (2010) *Real estate modelling and forecasting*. Cambridge: Cambridge University Press.

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

Brown, S. (1992) 'The wheel of retail gravitation?,' *Environment and Planning A: Economy and Space*, 24(10), pp. 1409–1429. https://doi.org/10.1068/a241409.

Buitelaar, E., N. Sorel, F. Verwest, F. van Dongen & A. Bregman (2013) *Gebiedsontwikkeling en commerciële vastgoedmarkten: een institutionele analyse van het (over)aanbod van winkels en kantoren*. Den Haag / Amsterdam: Planbureau voor de Leefomgeving / Amsterdam School of Real Estate.

Campo, K. et al. (2000) 'The impact of location factors on the attractiveness and optimal space shares of product categories,' *International Journal of Research in Marketing*, 17(4), pp. 255–279. https://doi.org/10.1016/s0167-8116(00)00026-4.

Centraal Bureau voor de Statistiek (2017) '65-plusser besteedt naar verhouding groot deel inkomen,' *Centraal Bureau Voor De Statistiek*, 17 March. https://www.cbs.nl/nl-nl/nieuws/2017/11/65-plusser-besteedt-naar-verhouding-groot-deel-inkomen.

Centraal Bureau voor de Statistiek (2023) 'Minder winkels in 2023,' *Centraal Bureau Voor De Statistiek*, 9 May. https://www.cbs.nl/nl-nl/nieuws/2023/19/minder-winkels-in-2023.

Centraal Bureau voor de Statistiek (no date) ‘Landelijk gebied,’ *Centraal Bureau Voor De Statistiek*, 2 May. https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/landelijk-gebied#:~:text=Telt%20de%20oad%20minder%20dan,er%20sprake%20van%20landelijk%20gebied.

Colwell, P.F. (2002) ‘Tweaking the DiPasquale-Wheaton Model,’ *Journal of Housing Economics*, 11, pp. 24-39.

Custers, A. (2022) ‘Staan grote winkelpanden vaker leeg?,*’ Locatus*, 28 April. https://locatus.com/blog/staan-grote-winkelpanden-vaker-leeg/#:~:text=Grootte%20pand%20versus%20kans%20op,daarom%20niet%20zomaar%20te%20trekken.

Custers, A. (2023) ‘Winkelleegstand Nederland daalt naar laagste niveau in meer dan 10 jaar,’ *Locatus*, 22 April. https://locatus.com/blog/winkelleegstand-nederland-daalt-naar-laagste-niveau-in-meer-dan-10-jaar/.

Dela Cruz, A. (2023) ‘Income-Expenditure Model | Definition, Equilibrium & Formula,’ *Study.com*, 2 February. https://study.com/academy/lesson/income-expenditure-model-overview-formula-examples.html.

Delage, M. et al. (2020) 'Retail decline in France’s small and medium-sized cities over four decades. Evidences from a multi-level analysis,' *Cities*, 104, p. 102790. https://doi.org/10.1016/j.cities.2020.102790.

DiPasquale, D. & Wheaton, W.C. (1992) ‘The Markets for Real Estate Assets and Space: A Conceptual Framework,’ *Journal of the American Real Estate and Urban Economics Association*, 20(1), pp. 191- 197.

Dolega, L. and Lord, A. (2020) 'Exploring the geography of retail success and decline: A case study of the Liverpool City Region,' *Cities*, 96, p. 102456. https://doi.org/10.1016/j.cities.2019.102456.

Evers, D., Tennekes, J. & van Dongen, F. (2014) *De bestendige binnenstad. Een verkenning van de leegstand, het leefklimaat, het voorzieningenniveau, het vestigingsklimaat en de economische structuur van de Nederlandse binnenstad*. Den Haag: Planbureau voor de leefomgeving.

Gool, P. van, D. Brounen, P. Jager & R.M. Weisz (2007) *Onroerend goed als belegging*. Groningen/Houten: Wolters-Noordhoff.

Hansson, L., Holmberg, U. and Post, A. (2022) 'Reorganising grocery shopping practices – the case of elderly consumers,' *The International Review of Retail, Distribution and Consumer Research*, 32(4), pp. 351–369. https://doi.org/10.1080/09593969.2022.2085137.

Hoekstra, J. & Vakili-Zad, C. (2011) ‘High Vacancy Rates and Rising House Prices: The Spanish Paradox,’ *Journal of Economic and Human Geography*, 102(1), pp. 55-71.

Hollander, J.B., Hart, M.D., Wiley, A. & Vavra, S. (2018) ‘Vacancy in Shrinking Downtowns: a Comparative Study of Québec, Ontario, and New England,’ *Journal of Housing and the Built Environment*, 33(4), pp. 591-613.

Ishikawa, T. and Toda, M. (2000) 'Some economic extensions of central-place theory involving profit maximisation,' *Urban Studies*, 37(3), pp. 481–495. https://doi.org/10.1080/0042098002078.

Kang, C.-D. (2019) 'Effect of neighborhood income and consumption on retail viability: Evidence from Seoul, Korea,' *Habitat International*, 94, pp. 102060. https://doi.org/10.1016/j.habitatint.2019.102060.

Kickert, C. (2021) ‘What’s in Store: Prospects and Challenges for American Street-Level commerce,’ *Journal of Urban Design*, 26(2), pp. 159-177.

Lagerwaard, V. (2022) ‘Transformatie winkels verbetert leefbaarheid in binnensteden,’ *Annexum*, 2 January. https://www.annexum.nl/nieuws/transformatie-winkels-verbetert-leefbaarheid-in-binnensteden/.

Lesáková, D. (2016) 'Seniors and their food shopping behavior: An Empirical analysis,' *Procedia - Social and Behavioral Sciences*, 220, pp. 243–250. https://doi.org/10.1016/j.sbspro.2016.05.496

Mallach, A. (2018) *The Empty House Next Door. Understanding and Reducing Vacancy and Hypervacancy in the United States*. Policy Focus Report. Cambridge, Massachussetts: Lincoln Insitute of Land Policy.

Meltzer, R. and Capperis, S. (2016) 'Neighbourhood differences in retail turnover: Evidence from New York City,' *Urban Studies*, 54(13), pp. 3022–3057. https://doi.org/10.1177/0042098016661268.

Ministerie van Algemene Zaken (2024) ‘Energielabel utiliteitsgebouwen,’ *Ministerie van algemene Zaken*, 28 May. https://www.rijksoverheid.nl/onderwerpen/energielabel-woningen-en-gebouwen/energielabel-utiliteitsbouw.

Ministerie van Algemene Zaken (2024a) ‘Coronavirus tijdlijn,’ *Ministerie van Algemene Zaken*, 4 June. https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdlijn.

Ministerie van Infrastructuur en Waterstaat (2022) ‘Corona geeft impuls aan online aankopen,’ *Ministerie van Infrastructuur en Waterstaat,* 15 May. https://www.kimnet.nl/actueel/nieuws/2022/03/24/corona-geeft-impuls-aan-online-aankopen.

Montgomery, D.C., Peck, E.A. & Vining, G.G. (2012) *Introduction to Linear Regression Analysis.* 6th edition. Hoboken: John Wiley & Sons, Inc.

Morgan, G.T. et al. (2022) 'Wellbeing as an emergent property of social practice,' *Buildings & Cities*, 3(1), p. 756. https://doi.org/10.5334/bc.262.

NOS (2022) 'Winkelstraten steeds leger: ‘Met alleen verkopen, verkopen, verkopen overleef je niet,’ *NOS*, 24 March. https://nos.nl/regio/noord-holland/artikel/248287-winkelstraten-steeds-leger-met-alleen-verkopen-verkopen-verkopen-overleef-je-niet.

Oskam, M. (2021) *Drivers of Retail Vacancy Pre-Covid19: The Effect of Building and Location Factors.* Groningen: Rijksuniversiteit Groningen.

PBL (2022) *Winkelleegstand in binnensteden voor, tijdens en na de pandemie*. https://www.pbl.nl/publicaties/winkelleegstand-in-binnensteden-voor-tijdens-en-na-de-pandemie.

Peek, G. & Veghel, G. V. (2011) *Geen stad zonder winkels*. In: Compacte Stad Extended: Rotterdam, 010 Publishers

Rabianski, J.S. (2002) ‘Vacancy in Market Analysis and Valuation,*’ Appraisal Journal*, 70(2), pp. 191-199.

Read, C. (1988) 'Advertising and natural vacancies in rental housing markets,' *Real Estate Economics*, 16(4), pp. 354–363. https://doi.org/10.1111/1540-6229.00460.

Reckwitz, A. (2002) ‘Toward A Theory of Social Practices: A Development in Culturalist Theorizing,’ *European Journal of Social Theory*, 5(2), pp. 243–263. doi:10.1177/13684310222225432.

Reilly, W. J. (1931) *The Law of Retail Gravitation*. New York: William J. Reilly.

Reimers, V. Clulow V. (2004) ‘Retail concentration: a comparison of spatial convenience in shopping strips and shopping centres,’ *Journal of Retailing and Consumer services*, 11 (4), pp 207-221.

Remøy, H.T., Koppels, P., Oel, C. van & Jonge, H. de (2007) *Characteristics of Vacant Offices, a DelphiApproach*. ENHR International Conference.

Schröder, P. (2023) ‘Hoe de retailbranche kan inspelen op de vergrijzing,’ *Vastgoedmarkt*, 2 February. https://www.vastgoedmarkt.nl/187161/hoe-de-retailbranche-kan-inspelen-op-de-vergrijzing.

Shove, E., M. Pantzar, and M. Watson (2012) *The Dynamics of Social Practice: Everyday Life and How It Changes*. London: Sage.

Slob, G. (2023) ‘Winkelleegstand Nederland daalt naar laagste niveau in meer dan 10 jaar’, *Locatus*, 2 February. https://locatus.com/blog/winkelleegstand-nederland-daalt-naar-laagste-niveau-in-meer-dan-10-jaar/.

Talen, E. and Park, J. (2021) 'Understanding urban retail vacancy,' *Urban Affairs Review*, 58(5), pp. 1411–1437. https://doi.org/10.1177/10780874211025451.

Thrift, N. and Kitchin, R. (2010) 'International Encyclopedia of Human Geography (2009),' *Choice (Chicago, Ill.)*, 48(01), pp. 48–0052. https://doi.org/10.5860/choice.48-0052.

Voordt, D.J.M. van der (2007) *Transformatie van kantoorgebouwen: thema’s, actoren, instrumenten en projecten*. Rotterdam: Uitgeverij 010.

Wang, X. and Hofe, R. (2008) *Research methods in urban and regional planning*. Springer Science & Business Media.

Warde, A. (2005) “Consumption and Theories of Practice.” *Journal of Consumer Culture,* 5(2), pp. 131–153. doi:10.1177/1469540505053090.

Warde, A. (2014) “After Taste: Culture, Consumption and Theories of Practice.” *Journal of Consumer Culture*, 14 (3), pp. 279–303. doi:10.1177/1469540514547828.

Yakubu, I.E. et al. (2017) 'Why are Older Inner-city Buildings Vacant? Implications for Town Centre Regeneration,' *Journal of Urban Regeneration and Renewal* [Preprint]. https://hstalks.com/article/906/why-are-older-inner-city-buildings-vacant-implicat/.

Zhang, D., Zhu, P. and Ye, Y. (2016) 'The effects of E-commerce on the demand for commercial real estate,' *Cities*, 51, pp. 106–120. https://doi.org/10.1016/j.cities.2015.11.012.

APPENDIX

Appendix A. Descriptive statistics (continued)

Table 1: Correlation matrix between the dependent variable and the independent variables

	lee~0inw	a_inw	a_00_14	a_15_24	a_25_44	a_45_64	a_65_oo
leeg_pe~0inw	1.0000						
a_inw	0.0877	1.0000					
a_00_14	0.0272	0.9696	1.0000				
a_15_24	0.0919	0.9490	0.8891	1.0000			
a_25_44	0.0944	0.9641	0.9240	0.9455	1.0000		
a_45_64	0.0756	0.9855	0.9629	0.9069	0.9168	1.0000	
a_65_oo	0.1269	0.9261	0.8735	0.8254	0.8180	0.9379	1.0000
g_hh_sti	-0.1493	-0.1581	-0.1151	-0.1909	-0.1839	-0.1279	-0.1381
gem_wvo	0.0707	-0.0642	-0.0513	-0.0562	-0.0563	-0.0657	-0.0765
gem_bouwjaar	-0.0701	0.0216	0.0702	-0.0083	-0.0071	0.0382	0.0182
stedelijkh~y	0.1767	0.4757	0.4436	0.4462	0.4810	0.4543	0.4461

	g_hh_sti	gem_wvo	gem_bo~r	stedel~y
g_hh_sti	1.0000			
gem_wvo	-0.0032	1.0000		
gem_bouwjaar	-0.0017	0.1830	1.0000	
stedelijkh~y	-0.2074	-0.0161	0.0637	1.0000

Table 2: Amount of zeros (districts with zero vacancy units per 1000 inhabitants) for the Y.

leeg_per_1k _inw	Freq.	Percent	Cum.
0	8,206	100.00	100.00
Total	8,206	100.00	

Appendix B. GIS maps

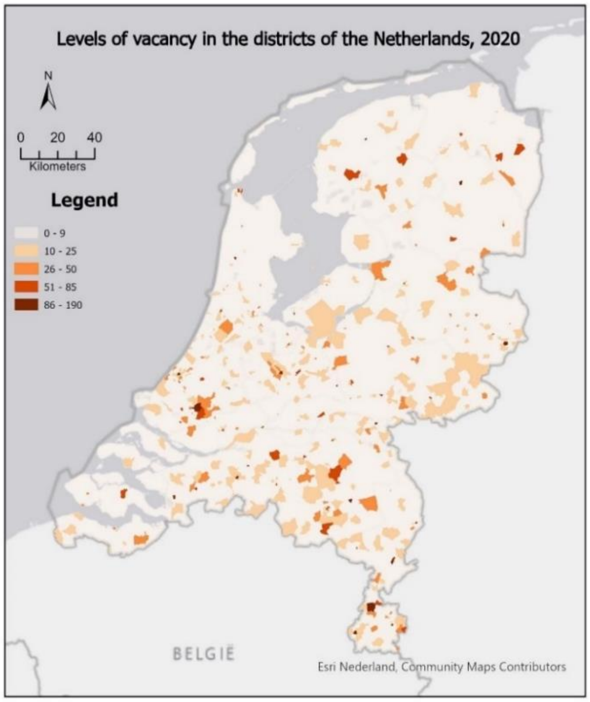

Figure 1: Map of the levels of vacancy in the districts of the Netherlands, 2020

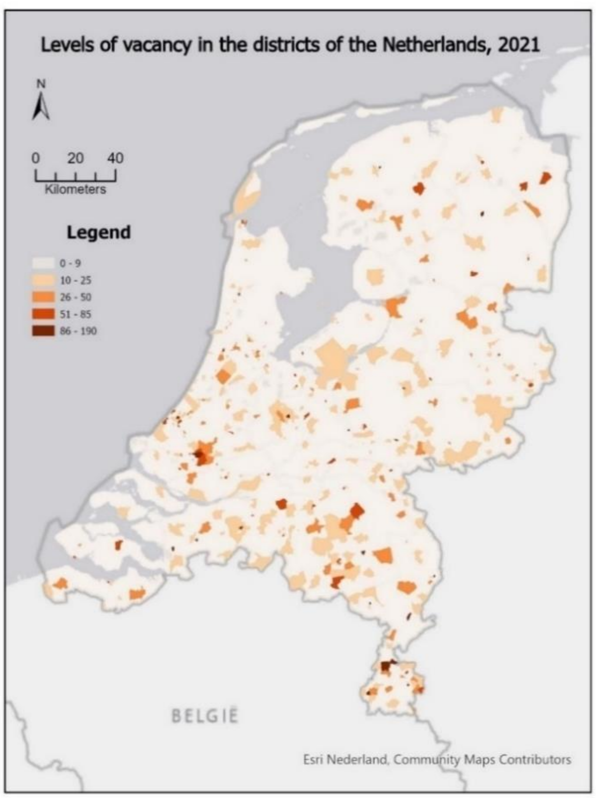

Figure 2: Map of the levels of vacancy in the districts of the Netherlands, 2021

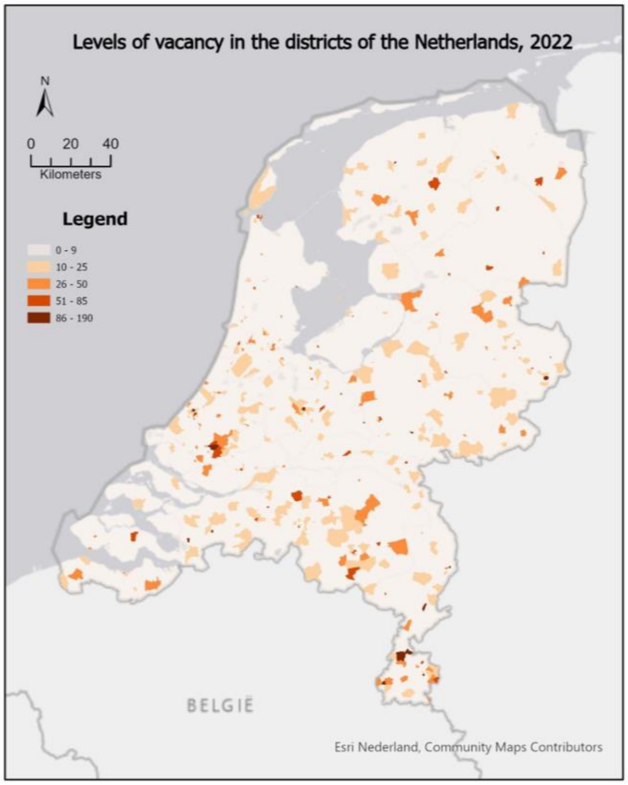

Figure 3: Map of the levels of vacancy in the districts of the Netherlands, 2022

Appendix C. Multiple linear regression assumptions

Table 3: Testing the multicollinearity assumption using VIF-scores

. vif

Variable	VIF	1/VIF
provincie~12	11.84	0.084444
provincie_~7	11.28	0.088677
provincie_~4	9.04	0.110611
provincie_~8	8.49	0.117742
provincie_~6	7.70	0.129942
provincie_~9	6.10	0.163908
provincie~10	5.52	0.181017
provincie_~3	4.11	0.243200
provincie~11	3.75	0.266602
provincie_~5	3.70	0.269999
provincie~y1	3.32	0.301513
a_25_44_sh~e	2.31	0.432627
a_45_64_sh~e	2.16	0.462152
dummy_2020	2.05	0.486620
dummy_2021	1.88	0.530773
a_00_14_sh~e	1.86	0.538123
g_hh_sti	1.78	0.561765
stedelijkh~y	1.74	0.575641
a_15_24_sh~e	1.60	0.625610
a_65_oo_sh~e	1.59	0.628677
ln_ainw	1.49	0.669799
gem_bouwjaar	1.12	0.891609
ln_gemwvo	1.06	0.946622
Mean VIF	4.15	

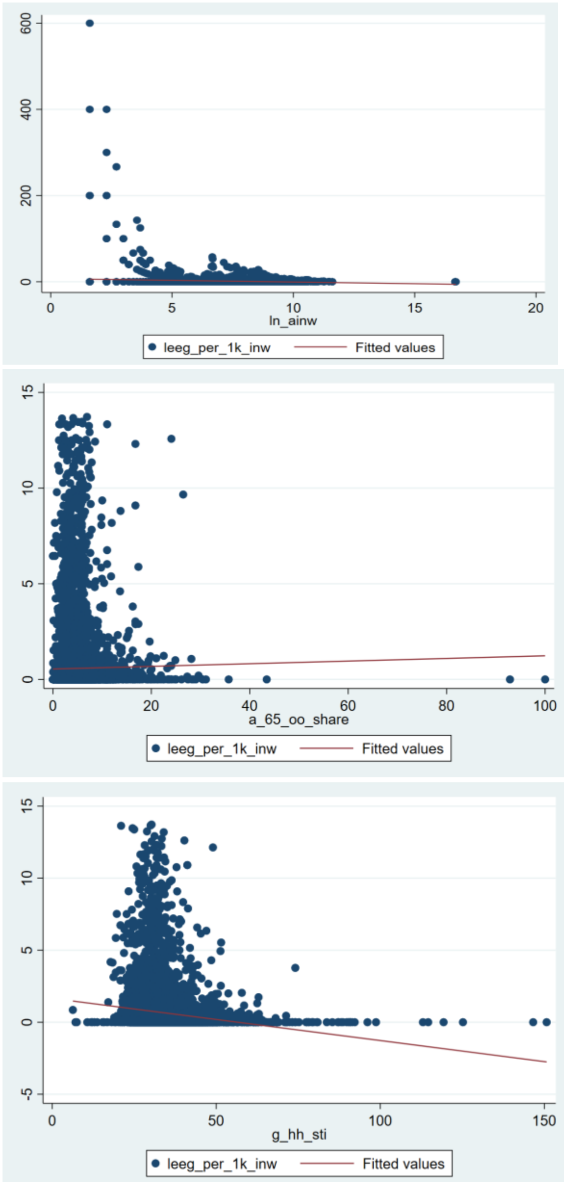

Figure 4: Testing the linearity assumption using a two-way scatterplot

Table 4: Testing the normality assumption using the Shapiro-Wilk test

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
leeg_pe~_inw	15,994	0.55782	3283.592	21.936	0.00000
ln_ainw	15,994	0.95684	320.505	15.632	0.00000
a_00_14_sh~e	15,994	0.59307	3021.823	21.711	0.00000
a_15_24_sh~e	15,994	0.40382	4427.189	22.745	0.00000
a_25_44_sh~e	15,994	0.43851	4169.592	22.583	0.00000
a_45_64_sh~e	15,994	0.54167	3403.518	22.033	0.00000
a_65_oo_sh~e	15,994	0.63319	2723.913	21.430	0.00000
g_hh_sti	9,759	0.80148	967.820	18.393	0.00000
ln_gemwvo	7,793	0.98137	74.672	11.467	0.00000
gem_bouwjaar	6,226	0.85947	462.514	16.210	0.00000
stedelijkh~y	15,994	0.99994	0.438	-2.237	0.98737

Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.

Table 5: Heteroskedasticity assumption using the Breusch-Pagan/Cook-Weisberg test

```
Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variables: ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share a_65_oo_share g_hh_sti
           ln_gemwvo gem_bouwjaar stedelijkheid_dummy dummy_2019 dummy_2020 dummy_2021 dummy_2022
           dummy_2023 provincie_dummy1 provincie_dummy2 provincie_dummy3 provincie_dummy4
           provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8 provincie_dummy9
           provincie_dummy10 provincie_dummy11 provincie_dummy12

H0: Constant variance

   chi2(23) = 4231.27
Prob > chi2 =  0.0000
```

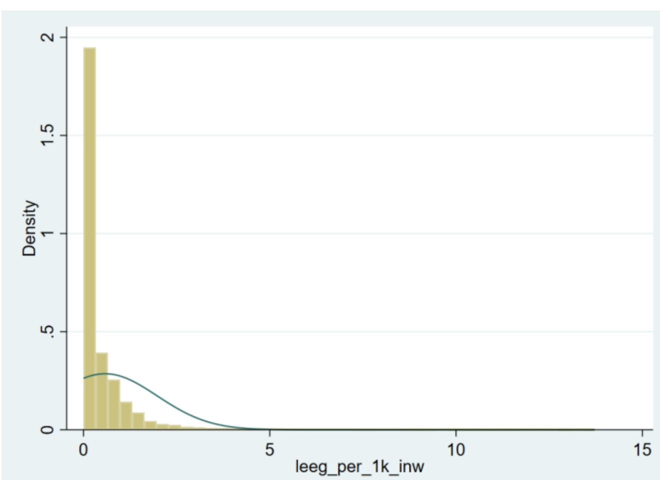

Figure 5: Histogram dependent variable

Appendix D. Regression models (continued)

Table 6: Regression models 6 and 7 with control variables building year and WVO

Variables	(6)	(7)
ln_ainw	-0.342***	-0.341***
	(0.032)	(0.039)
a_00_14_share	-0.992***	-1.018***
	(0.054)	(0.062)
a_15_24_share	-0.110**	-0.103*
	(0.054)	(0.060)
a_25_44_share	0.421***	0.423***
	(0.045)	(0.052)
a_45_64_share	-0.091***	-0.084**
	(0.032)	(0.039)
a_65_oo_share	0.245***	0.262***
	(0.026)	(0.031)
dummy_2020	-0.053	
	(0.052)	
dummy_2021	-0.131**	-0.075
	(0.054)	(0.054)
dummy_2022	-0.260***	-0.217***
	(0.065)	(0.066)
stedelijkheid_dummy	0.404***	0.435***
	(0.048)	(0.057)
provincie_dummy1	-0.580***	-0.640***
	(0.166)	(0.203)
provincie_dummy3	-0.589***	-0.821***
	(0.134)	(0.152)
provincie_dummy4	-0.638***	-0.777***
	(0.124)	(0.143)
provincie_dummy5	-0.777***	-0.938***
	(0.149)	(0.173)
provincie_dummy6	-0.781***	-0.877***
	(0.136)	(0.159)
provincie_dummy7	-0.718***	-0.804***
	(0.122)	(0.140)
provincie_dummy8	-0.802***	-0.888***
	(0.132)	(0.154)
provincie_dummy9	-0.418***	-0.551***
	(0.128)	(0.146)
provincie_dummy10	-0.485***	-0.574***
	(0.140)	(0.164)
provincie_dummy11	-0.315**	-0.462**
	(0.160)	(0.187)
provincie_dummy12	-0.734***	-0.826***
	(0.124)	(0.143)
g_hh_sti	0.026***	0.028***
	(0.006)	(0.007)
ln_gemwvo	0.245***	0.298***
	(0.026)	(0.030)
gem_bouwjaar		-0.003***
		(0.001)
Constant	3.352***	9.382***
	(0.431)	(1.367)
Observations	5,456	3,930
R-squared	0.289	0.307

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix E. Sensitivity analysis

$$F = \frac{(\mathrm{RSS}_{\mathrm{pooled}} - (\mathrm{RSS}_1 + \mathrm{RSS}_2))/k}{(\mathrm{RSS}_1 + \mathrm{RSS}_2)/(n_1 + n_2 - 2k)}$$

Figure 6: Formula used for the Chow F-test on urbanity level

$$\mathrm{F} = \frac{(24165.920-(6373.163+16607.448))/22}{(6373.163+16607.448)\ /\ (9785+6209-2\ x\ 22)} \sim \mathrm{F}[22, 15950]$$

Figure 7: Conducted Chow F-test on urbanity level

Table 7: Pooled model with interaction terms for the Chow F-test on urbanity level

VARIABLES	Pooled model with interaction terms leeg_per_1k_inw
ln_ainw	-0.170***
	(0.010)
1.stedelijkheid_dummy	0.897***
	(0.149)
0b.stedelijkheid_dummy#co.ln_ainw	0.000
	(0.000)
1.stedelijkheid_dummy#c.ln_ainw	-0.049***
	(0.017)
a_00_14_share	-0.022*
	(0.011)
0b.stedelijkheid_dummy#co.a_00_14_share	0.000
	(0.000)
1.stedelijkheid_dummy#c.a_00_14_share	-0.198***
	(0.016)
a_15_24_share	0.006
	(0.008)
0b.stedelijkheid_dummy#co.a_15_24_share	0.000
	(0.000)
1.stedelijkheid_dummy#c.a_15_24_share	0.036***
	(0.013)
a_25_44_share	0.008
	(0.008)
0b.stedelijkheid_dummy#co.a_25_44_share	0.000
	(0.000)
1.stedelijkheid_dummy#c.a_25_44_share	0.034***
	(0.009)
a_45_64_share	-0.003
	(0.006)
0b.stedelijkheid_dummy#co.a_45_64_share	0.000
	(0.000)
1.stedelijkheid_dummy#c.a_45_64_share	-0.040***
	(0.012)
a_65_oo_share	-0.007
	(0.006)
0b.stedelijkheid_dummy#co.a_65_oo_share	0.000
	(0.000)
1.stedelijkheid_dummy#c.a_65_oo_share	0.133***
	(0.010)

dummy_2020	0.020
	(0.031)
dummy_2021	-0.000
	(0.030)
dummy_2022	-0.037
	(0.030)
dummy_2023	-0.080***
	(0.030)
provincie_dummy1	1.407***
	(0.076)
provincie_dummy3	1.138***
	(0.066)
provincie_dummy4	1.162***
	(0.045)
provincie_dummy5	1.108***
	(0.073)
provincie_dummy6	1.403***
	(0.047)
provincie_dummy7	1.467***
	(0.038)
provincie_dummy8	1.134***
	(0.044)
provincie_dummy9	1.347***
	(0.052)
provincie_dummy10	1.419***
	(0.058)
provincie_dummy11	1.750***
	(0.067)
provincie_dummy12	1.156***
	(0.038)
Constant	1.202***
	(0.082)
Observations	15,994
R-squared	0.259

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
( 1)  1.stedelijkheid_dummy#c.ln_ainw = 0
( 2)  1.stedelijkheid_dummy#c.a_00_14_share = 0
( 3)  1.stedelijkheid_dummy#c.a_15_24_share = 0
( 4)  1.stedelijkheid_dummy#c.a_25_44_share = 0
( 5)  1.stedelijkheid_dummy#c.a_45_64_share = 0
( 6)  1.stedelijkheid_dummy#c.a_65_oo_share = 0

      F(  6, 15965) =   66.79
           Prob > F =    0.0000
```

Figure 8: Conducted Chow F-test with interaction terms on urbanity level in STATA

$$F = \frac{(\mathrm{RSS}_{\mathrm{pooled}} - (\mathrm{RSS}_{2019} + \mathrm{RSS}_{2020/2021} + \mathrm{RSS}_{2022/2023}))/k}{(\mathrm{RSS}_{2019} + \mathrm{RSS}_{2020/2021} + \mathrm{RSS}_{2022/2023})/(n_{2019} + n_{2020/2021} + n_{2022/2023} - 3k)}$$

Figure 9: Formula used for the Chow F-test on COVID-19

$$F = \frac{(24193.174-(4281.823+9993.060+9619.585))/18}{(4281.823+9993.060+9619.585)/\,(3103+6317+6574-3\ x\ 18)}\ \sim\ F[18, 15940]$$

Figure 10: Conducted Chow F-test on COVID-19

Table 8: Pooled model with interaction terms for the Chow F-test on COVID-19

Variables	Pooled model with interaction terms leeg_per_1k_inw
ln_ainw	-0.131***
	(0.011)
1.dummy_2019	-0.012
	(0.163)
0b.dummy_2019#co.ln_ainw	0.000
	(0.000)
1.dummy_2019#c.ln_ainw	0.001
	(0.019)
1.dummy_2020_2021	0.280*
	(0.146)
0b.dummy_2020_2021#co.ln_ainw	0.000
	(0.000)
1.dummy_2020_2021#c.ln_ainw	-0.013
	(0.016)
o.ln_ainw	-
1o.dummy_2022_2023	-
0b.dummy_2022_2023#co.ln_ainw	0.000
	(0.000)
1o.dummy_2022_2023#co.ln_ainw	0.000
	(0.000)
a_00_14_share	-0.096***
	(0.010)
0b.dummy_2019#co.a_00_14_share	0.000
	(0.000)
1.dummy_2019#c.a_00_14_share	-0.031
	(0.020)
0b.dummy_2020_2021#co.a_00_14_share	0.000
	(0.000)
1.dummy_2020_2021#c.a_00_14_share	-0.175***
	(0.019)
o.a_00_14_share	-
0b.dummy_2022_2023#co.a_00_14_share	0.000
	(0.000)
1o.dummy_2022_2023#co.a_00_14_share	0.000
	(0.000)
a_15_24_share	0.027***
	(0.009)
0b.dummy_2019#co.a_15_24_share	0.000
	(0.000)
1.dummy_2019#c.a_15_24_share	-0.006
	(0.014)
0b.dummy_2020_2021#co.a_15_24_share	0.000
	(0.000)
1.dummy_2020_2021#c.a_15_24_share	0.018
	(0.020)
o.a_15_24_share	-
0b.dummy_2022_2023#co.a_15_24_share	0.000

	(0.000)
1o.dummy_2022_2023#co.a_15_24_share	0.000
	(0.000)
a_25_44_share	0.028***
	(0.004)
0b.dummy_2019#co.a_25_44_share	0.000
	(0.000)
1.dummy_2019#c.a_25_44_share	0.037***
	(0.011)
0b.dummy_2020_2021#co.a_25_44_share	0.000
	(0.000)
1.dummy_2020_2021#c.a_25_44_share	0.097***
	(0.011)
o.a_25_44_share	-
0b.dummy_2022_2023#co.a_25_44_share	0.000
	(0.000)
1o.dummy_2022_2023#co.a_25_44_share	0.000
	(0.000)
a_45_64_share	-0.021***
	(0.006)
0b.dummy_2019#co.a_45_64_share	0.000
	(0.000)
1.dummy_2019#c.a_45_64_share	-0.003
	(0.012)
0b.dummy_2020_2021#co.a_45_64_share	0.000
	(0.000)
1.dummy_2020_2021#c.a_45_64_share	-0.036***
	(0.013)
o.a_45_64_share	-
0b.dummy_2022_2023#co.a_45_64_share	0.000
	(0.000)
1o.dummy_2022_2023#co.a_45_64_share	0.000
	(0.000)
a_65_oo_share	0.033***
	(0.006)
0b.dummy_2019#co.a_65_oo_share	0.000
	(0.000)
1.dummy_2019#c.a_65_oo_share	0.007
	(0.013)
o.a_65_oo_share	-
0b.dummy_2020_2021#co.a_65_oo_share	0.000
	(0.000)
1.dummy_2020_2021#c.a_65_oo_share	0.038***
	(0.012)
0b.dummy_2022_2023#co.a_65_oo_share	0.000
	(0.000)
1o.dummy_2022_2023#co.a_65_oo_share	0.000
	(0.000)
provincie_dummy1	1.322***
	(0.078)
provincie_dummy3	1.084***
	(0.067)
provincie_dummy4	1.165***
	(0.045)
provincie_dummy5	1.066***
	(0.075)
provincie_dummy6	1.384***
	(0.048)
provincie_dummy7	1.481***
	(0.038)

provincie_dummy8	1.247***
	(0.045)
provincie_dummy9	1.364***
	(0.053)
provincie_dummy10	1.488***
	(0.059)
provincie_dummy11	1.664***
	(0.068)
provincie_dummy12	1.245***
	(0.039)
Constant	1.020***
	(0.092)
Observations	15,994
R-squared	0.233

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
( 1)  1.dummy_2019#c.ln_ainw = 0
( 2)  1.dummy_2020_2021#c.ln_ainw = 0
( 3)  1o.dummy_2022_2023#co.ln_ainw = 0
( 4)  1.dummy_2019#c.a_00_14_share = 0
( 5)  1.dummy_2020_2021#c.a_00_14_share = 0
( 6)  1o.dummy_2022_2023#co.a_00_14_share = 0
( 7)  1.dummy_2019#c.a_15_24_share = 0
( 8)  1.dummy_2020_2021#c.a_15_24_share = 0
( 9)  1o.dummy_2022_2023#co.a_15_24_share = 0
(10)  1.dummy_2019#c.a_25_44_share = 0
(11)  1.dummy_2020_2021#c.a_25_44_share = 0
(12)  1o.dummy_2022_2023#co.a_25_44_share = 0
(13)  1.dummy_2019#c.a_45_64_share = 0
(14)  1.dummy_2020_2021#c.a_45_64_share = 0
(15)  1o.dummy_2022_2023#co.a_45_64_share = 0
(16)  1.dummy_2019#c.a_65_oo_share = 0
(17)  1.dummy_2020_2021#c.a_65_oo_share = 0
(18)  1o.dummy_2022_2023#co.a_65_oo_share = 0

       F( 12, 15962) =   13.39
            Prob > F =    0.0000
```

Figure 11: Conducted Chow F-test with interaction terms on COVID-19 in STATA

Appendix F. STATA Syntax

```
*importing merged and partially cleaned dataset*
import excel "C:\Users\tessa\MyProject\Query2019-2023.xlsx", sheet("all") firstrow

summarize
browse
codebook
```

```
*destring of string variables*
destring a_inw a_man a_vrouw a_00_14 a_15_24 a_25_44 a_45_64 a_65_oo a_geb p_geb a_ste p_ste
a_hh a_1p_hh a_hh_z_k a_hh_m_k g_hhgro bev_dich a_inkont g_ink_po g_ink_pi g_hh_sti ste_mvs
ste_oad RRI RRI_PAND RRI_STRAAT RRI_BRANCHE RRI_MARKT RRI_COVID19, replace

replace g_hhgro = subinstr(g_hhgro, ",", ".", .)
destring g_hhgro, replace

replace g_ink_po = subinstr(g_ink_po, ",", ".", .)
destring g_ink_po, replace

replace g_ink_pi = subinstr(g_ink_pi, ",", ".", .)
destring g_ink_pi, replace

replace g_hh_sti = subinstr(g_hh_sti, ",", ".", .)
destring g_hh_sti, replace

*tostring of numeric variable*
tostring HUISNR, gen(HUISNR_string)
drop HUISNR

*more data cleaning*
drop GOADID
drop WOONPLAATS
drop KWBID
drop if recs == "Buurt"
drop if recs == "Gemeente"

*creating dummies for vacancy and rural vs urban*
gen leegstand_dummy = !missing(NAAM)
gen stedelijkheid_dummy = (ste_oad > 1000)

*creating the levels of vacancy per district per year*
tostring JAAR, replace
egen leegstand2019 = total(leegstand_dummy) if JAAR == "2019", by(gwb_code)

egen leegstand2020 = total(leegstand_dummy) if JAAR == "2020", by(gwb_code)

egen leegstand2021 = total(leegstand_dummy) if JAAR == "2021", by(gwb_code)

egen leegstand2022 = total(leegstand_dummy) if JAAR == "2022", by(gwb_code)

egen leegstand2023 = total(leegstand_dummy) if JAAR == "2023", by(gwb_code)
```

```
destring JAAR, replace

egen leegstandallejaren = rowtotal(leegstand2019 leegstand2020 leegstand2021 leegstand2022
leegstand2023)

*creating the average of the control variables per year*
egen gem_wvo = mean(WVO), by(gwb_code JAAR)
egen gem_bouwjaar = mean(BAG_BOUWJAAR), by(gwb_code JAAR)

*creating the levels of vacancy per year again and dropping NL level and duplicates*
egen leeg = total(leegstand_dummy), by(gwb_code JAAR)

egen tot = count(leegstand_dummy), by(gwb_code JAAR)

duplicates drop gwb_code JAAR, force

drop in 1/5

*Correlation matrix*
pwcorr leeg_per_1000inw a_inw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti gem_wvo gem_bouwjaar stedelijkheid_dummy

*checking if the vars are normally distributed and log-transformations*
histogram leeg, normal
histogram a_inw, normal
histogram a_00_14, normal
histogram a_15_24, normal
histogram a_25_44, normal
histogram a_45_64, normal
histogram a_65_oo, normal
histogram g_hh_sti, normal
histogram gem_wvo, normal
histogram gem_bouwjaar, normal

gen ln_ainw = ln(a_inw)
histogram ln_ainw, normal

gen ln_a0014 = ln(a_00_14)
histogram ln_a0014, normal

gen ln_a1524 = ln(a_15_24)
histogram ln_a1524, normal

gen ln_a2544 = ln(a_25_44)
```

```
histogram ln_a2544, normal

gen ln_a4564 = ln(a_45_64)
histogram ln_a4564, normal

gen ln_a65oo = ln(a_65_oo)
histogram ln_a65oo, normal

gen ln_gemwvo = ln(gem_wvo)
histogram ln_gemwvo, normal

*Creating new Y*
gen leeg_per_1k_inw = leeg/(a_inw/1000)

*converting age groups into shares*
egen totaal_inw_per_wijk = total(a_inw), by(gwb_code)

gen a_00_14_share = (a_00_14 / totaal_inw_per_wijk) * 100
gen a_15_24_share = (a_15_24 / totaal_inw_per_wijk) * 100
gen a_25_44_share = (a_25_44 / totaal_inw_per_wijk) * 100
gen a_45_64_share = (a_45_64 / totaal_inw_per_wijk) * 100
gen a_65_oo_share = (a_65_oo / totaal_inw_per_wijk) * 100

*checking*
tabulate leeg_per_1k_inw if leeg_per_1k_inw == 0
tabulate gem_bouwjaar if gem_bouwjaar == 0
tabulate ln_gemwvo if ln_gemwvo == 0

histogram leeg_per_1k_inw, normal

*creating location and time fixed effects*
gen dummy_2019 = (JAAR == 2019)
gen dummy_2020 = (JAAR == 2020)
gen dummy_2021 = (JAAR == 2021)
gen dummy_2022 = (JAAR == 2022)
gen dummy_2023 = (JAAR == 2023)

tabulate PROVINCIE, gen(provincie_dummy)

*checking missing values*
tabulate leeg_per_1k_inw if missing(leeg_per_1k_inw)
tabulate ln_ainw if missing(ln_ainw)
tabulate a_00_14_share if missing(a_00_14_share)
tabulate g_hh_sti if missing(g_hh_sti)
```

```
tabulate gem_bouwjaar if missing(gem_bouwjaar)
tabulate ln_gemwvo if missing(ln_gemwvo)

reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti ln_gemwvo gem_bouwjaar stedelijkheid_dummy dummy_2019 dummy_2020
dummy_2021 dummy_2022 dummy_2023 provincie_dummy1 provincie_dummy2 provincie_dummy3
provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8
provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12 if leegstand_dummy
== 1

reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti ln_gemwvo gem_bouwjaar stedelijkheid_dummy dummy_2019 dummy_2020
dummy_2021 dummy_2022 dummy_2023 provincie_dummy1 provincie_dummy2 provincie_dummy3
provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8
provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12 if leegstand_dummy
== 0

**checking assumptions**
*multicollinearity*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti ln_gemwvo gem_bouwjaar stedelijkheid_dummy dummy_2019 dummy_2020
dummy_2021 dummy_2022 dummy_2023 provincie_dummy1 provincie_dummy2 provincie_dummy3
provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8
provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
vif
*linearity*
twoway (scatter leeg_per_1k_inw ln_ainw) (lfit leeg_per_1k_inw ln_ainw)
twoway (scatter leeg_per_1k_inw a_00_14_share) (lfit leeg_per_1k_inw a_00_14_share)
twoway (scatter leeg_per_1k_inw a_65_oo_share) (lfit leeg_per_1k_inw a_65_oo_share)
twoway (scatter leeg_per_1k_inw g_hh_sti) (lfit leeg_per_1k_inw g_hh_sti)
*normality of residuals*
swilk leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti ln_gemwvo gem_bouwjaar stedelijkheid_dummy
dummy_2019 dummy_2020 dummy_2021 dummy_2022 dummy_2023 provincie_dummy1
provincie_dummy2 provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6
provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11
provincie_dummy12
*homoscedasticity*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti ln_gemwvo gem_bouwjaar stedelijkheid_dummy dummy_2019 dummy_2020
dummy_2021 dummy_2022 dummy_2023 provincie_dummy1 provincie_dummy2 provincie_dummy3
provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8
provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
```

```
hettest ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share a_65_oo_share g_hh_sti
ln_gemwvo gem_bouwjaar stedelijkheid_dummy dummy_2019 dummy_2020 dummy_2021
dummy_2022 dummy_2023 provincie_dummy1 provincie_dummy2 provincie_dummy3
provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8
provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12

*Testing other new Y option: changes in vacancy over time per district*
egen gwb_c=group(gwb_code)
xtset gwb_c JAAR
drop if gwb_code=="NL00"
sort gwb_c JAAR
bysort gwb_c: gen delta_leegstand = leeg - L.leeg
order leeg delta_leegstand
browse

*The dependent variable and removing outliers*
gen leeg_per_1000inw=leeg/a_inw*1000
sum
graph box leeg_per_1000inw if leeg_per_1000inw~=0
graph box leeg_per_1000inw if leeg_per_1000inw~=0 & leeg_per_1000inw<10

summarize leeg_per_1000inw, detail
local cutoff = r(p99)
drop if leeg_per_1000inw > `cutoff'

*Model building with change Y*
reg delta_leegstand ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share, robust

**Model building Y 1000inh. **
ssc install outreg 2, replace
*model 1*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share, robust
outreg2 using mythesisoutput2, word dec(3) replace ctitle(model 1)
*model 2*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2020 dummy_2021 dummy_2022 dummy_2023, robust
outreg2 using mythesisoutput2, word dec(3) append ctitle(model 2)
*model 3*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2020 dummy_2021 dummy_2022 dummy_2023 stedelijkheid_dummy, robust
outreg2 using mythesisoutput2, word dec(3) append ctitle(model 3)
*model 4*
```

```
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share stedelijkheid_dummy dummy_2020 dummy_2021 dummy_2022 dummy_2023
provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6
provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11
provincie_dummy12, robust
outreg2 using mythesisoutput2, word dec(3) append ctitle(model 4)
*model 5*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share g_hh_sti stedelijkheid_dummy dummy_2020 dummy_2021 dummy_2022
provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6
provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11
provincie_dummy12 if inlist(JAAR, 2019, 2020, 2021, 2022), robust
outreg2 using mythesisoutput2, word dec(3) append ctitle(model 5)
*model 6*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2020 dummy_2021 dummy_2022 stedelijkheid_dummy provincie_dummy1
provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7
provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
g_hh_sti ln_gemwvo if inlist(JAAR, 2019, 2020, 2021, 2022), robust
*model 7*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2021 dummy_2022 stedelijkheid_dummy provincie_dummy1
provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7
provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
g_hh_sti ln_gemwvo gem_bouwjaar if inlist(JAAR, 2020, 2021, 2022), robust

**Chow test urbanity level**
*Unrestricted models*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2020 dummy_2021 dummy_2022 dummy_2023 provincie_dummy1
provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7
provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
if stedelijkheid_dummy == 0
estimates store urban0

reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2020 dummy_2021 dummy_2022 dummy_2023 provincie_dummy1
provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7
provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
if stedelijkheid_dummy == 1
estimates store urban1
```

```
*pooled model without interaction terms*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share dummy_2020 dummy_2021 dummy_2022 dummy_2023 provincie_dummy1
provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7
provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12

*pooled model with interaction terms*
reg leeg_per_1k_inw c.ln_ainw##i.stedelijkheid_dummy c.a_00_14_share##i.stedelijkheid_dummy
c.a_15_24_share##i.stedelijkheid_dummy c.a_25_44_share##i.stedelijkheid_dummy
c.a_45_64_share##i.stedelijkheid_dummy c.a_65_oo_share##i.stedelijkheid_dummy dummy_2020
dummy_2021 dummy_2022 dummy_2023 provincie_dummy1  provincie_dummy3 provincie_dummy4
provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8 provincie_dummy9
provincie_dummy10 provincie_dummy11 provincie_dummy12
estimates store combined

*Test*
test 1.stedelijkheid_dummy#c.ln_ainw 1.stedelijkheid_dummy#c.a_00_14_share
1.stedelijkheid_dummy#c.a_15_24_share 1.stedelijkheid_dummy#c.a_25_44_share
1.stedelijkheid_dummy#c.a_45_64_share 1.stedelijkheid_dummy#c.a_65_oo_share

**Chow test years**
*new dummies*
gen dummy_2020_2021 = (JAAR == 2020 | JAAR == 2021)
gen dummy_2022_2023 = (JAAR == 2022 | JAAR == 2023)

*Unrestricted models*
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5
provincie_dummy6 provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10
provincie_dummy11 provincie_dummy12 if dummy_2019 == 1
estimates store year_2019

reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5
provincie_dummy6 provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10
provincie_dummy11 provincie_dummy12 if dummy_2020 == 1 | dummy_2021 == 1
estimates store year_2020_2021

reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share
a_65_oo_share provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5
provincie_dummy6 provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10
provincie_dummy11 provincie_dummy12 if dummy_2022 == 1 | dummy_2023 == 1
estimates store year_2022_2023
```

Pooled model without interaction terms
reg leeg_per_1k_inw ln_ainw a_00_14_share a_15_24_share a_25_44_share a_45_64_share a_65_oo_share provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12

Pooled model with interaction terms
reg leeg_per_1k_inw c.ln_ainw##i.dummy_2019 c.ln_ainw##i.dummy_2020_2021 c.ln_ainw##i.dummy_2022_2023 ///
c.a_00_14_share##i.dummy_2019 c.a_00_14_share##i.dummy_2020_2021 c.a_00_14_share##i.dummy_2022_2023 ///
c.a_15_24_share##i.dummy_2019 c.a_15_24_share##i.dummy_2020_2021 c.a_15_24_share##i.dummy_2022_2023 ///
c.a_25_44_share##i.dummy_2019 c.a_25_44_share##i.dummy_2020_2021 c.a_25_44_share##i.dummy_2022_2023 ///
c.a_45_64_share##i.dummy_2019 c.a_45_64_share##i.dummy_2020_2021 c.a_45_64_share##i.dummy_2022_2023 ///
c.a_65_oo_share##i.dummy_2019 c.a_65_oo_share##i.dummy_2020_2021 c.a_65_oo_share##i.dummy_2022_2023 ///
provincie_dummy1 provincie_dummy3 provincie_dummy4 provincie_dummy5 provincie_dummy6 provincie_dummy7 ///
provincie_dummy8 provincie_dummy9 provincie_dummy10 provincie_dummy11 provincie_dummy12
estimates store combined

test
test 1.dummy_2019#c.ln_ainw 1.dummy_2020_2021#c.ln_ainw 1.dummy_2022_2023#c.ln_ainw ///
1.dummy_2019#c.a_00_14_share 1.dummy_2020_2021#c.a_00_14_share 1.dummy_2022_2023#c.a_00_14_share ///
1.dummy_2019#c.a_15_24_share 1.dummy_2020_2021#c.a_15_24_share 1.dummy_2022_2023#c.a_15_24_share ///
1.dummy_2019#c.a_25_44_share 1.dummy_2020_2021#c.a_25_44_share 1.dummy_2022_2023#c.a_25_44_share ///
1.dummy_2019#c.a_45_64_share 1.dummy_2020_2021#c.a_45_64_share 1.dummy_2022_2023#c.a_45_64_share ///
1.dummy_2019#c.a_65_oo_share 1.dummy_2020_2021#c.a_65_oo_share 1.dummy_2022_2023#c.a_65_oo_share