university of groningen

faculty of spatial sciences

"Living in diverse neighbourhoods"

The association between employment sectors and house prices in Amsterdam

Jordi van den Brink

August 2021

COLOFON

Title	:	Living in diverse neighbourhoods - the association between employment sectors and house prices in Amsterdam
Version	:	Final
Author	:	Jordi van den Brink
University	:	University of Groningen
Faculty	:	Spatial Sciences
Study	:	Master Real Estate Studies
Student number	:	S3553337
E-mail	:	Jordi.vdBrink@outlook.com
Date	:	20 August 2021
	:	
Supervisor	:	Prof. dr. E. (Ed) Nozeman
Assessor	:	Prof. dr. ir. A.J. (Arno) van der Vlist

Disclaimer: “Master theses are preliminary materials to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the author and do not indicate concurrence by the supervisor or research staff.”

ABSTRACT

The aim of this research is to investigate to what extent a greater variety of employment is associated with higher property prices. The scale of this study applies to neighbourhoods in Amsterdam and serve as a case study. The adopted methodology involves a multiple linear regression model and the variety of employment within residential neighbourhoods is measured using a diversity index. The presence of sectors ´government, education, and healthcare´ and ´culture and recreation´ are positively associated with property prices in the neighbourhood, whereas the presence of sectors ´construction´ and ´trade and catering´ are negatively associated with property prices. The result is that 'the specific neighbourhoods with more diversity' are associated with higher property prices, although the association is not strong. On the other hand, an increase in the variety of employment sectors in surrounding neighbourhoods is strongly associated with higher property prices. Neighbourhoods with a lower population density are stronger associated with higher property values when having more diverse surrounding neighbourhoods, compared with neighbourhoods with middle or higher population densities in Amsterdam. This study empirically indicates that neighbourhood property prices do not solely depend on the characteristics within the neighbourhood but also on the composition of surrounding neighbourhoods.

Keywords: residential property price, neighbourhood, mixed use, diversity index, employment type, linear regression model, ordinary least squares

ACKNOWLEDGEMENTS

This thesis is written for the finalization of the master Real Estate Studies at the University of Groningen. The master is a valuable addition to my bachelor Construction Management.

I wish to thank the people who made writing this thesis possible. First, I would like to address my appreciation to my supervisor prof. dr. Ed Nozeman. I have experienced the supervision as very pleasant and I am extremely grateful for the thorough feedback, the interesting discussions, and his expertise on real estate. In addition, I appreciated the clear communication and quick responses. His knowledge and skills have been a true inspiration for me, thank you. Also, I want to thank dr. Sierd-Jan Koster for providing the LISA-dataset which was of great importance for the results in this research. Thereby, I want to express my appreciation to dr. Mark van Duijn for providing valuable feedback on my thesis proposal. The Geodienst of the University of Groningen, which provided excellent help with technical concerns related to the program GIS (Geographic Information System), was also a valuable service for me. Furthermore, I want to express my gratitude to the assessor, prof. dr. ir. Arno van der Vlist, for giving in-depth feedback and taking the time and effort to review this thesis. Finally, I am thankful for my family and friends who have supported me throughout the period of writing this thesis.

Jordi van den Brink

TABLE OF CONTENTS

1. INTRODUCTION ... 1
 1.1 Motivation ... 1
 1.2 Literature review ... 3
 1.3 Problem statement, aim, and research questions ... 4
 1.4 Conceptual model ... 5
 1.5 Methodology and data ... 5
 1.6 Thesis overview ... 6

2. THEORETICAL FRAMEWORK ... 7
 2.1 Mixed-use in theory ... 7
 2.2 Values of properties ... 9
 2.3 Mixed use and property prices ... 10
 2.4 Hypotheses ... 12

3. DATA AND METHOD ... 14
 3.1 Data ... 14
 3.2 Diversity index ... 16
 3.3 Linear regression models ... 19
 3.4 Descriptive statistics ... 22
 3.5 Sensitivity analysis ... 23

4. RESULTS ... 24
 4.1 Hypothesis 1 – 4 ... 25
 4.5 Other variables of interest ... 32
 4.6 Robustness test 1 – 4 ... 34

5. CONCLUSION AND FUTURE RESEARCH ... 47
 5.1 Conclusion and discussion ... 47
 5.2 Policy relevance ... 51
 5.3 Limitations and future research ... 51

REFERENCES ... 53

APPENDICES ... 58

1. INTRODUCTION

1.1 Motivation

From the dawn of humanity, more than a million years ago, humans have lived in small and intimate communities, where people's natural environment has focused on multifunctionality (Harari 2014, p. 398). Since the Agricultural Revolution 12,000 years ago, people generally have lived in close proximity to their place of work and transport has mainly been by foot, creating a relatively dense area where people work and live. Due to the large growth in population, cities expanded and transportation became motorised. Especially after World War II, the number of cars sharply increased, leading to increased commuting distances, decentralisation, and a more segregated form of living (Nabil and Eldayem, 2014). Many cars not only lead to traffic congestion but also negatively impact biodiversity and people's quality of life, which translates into loss of time as well as economic loss (Gössling, 2020). Furthermore, cities underwent a transition and became more specialised, and they were divided into zones for working and living (Vorontsova et al., 2016). Consequently, the idea of 'combining working and living in smaller areas', also referred to as mixed use, faded and was no longer applied after World War II according to Nabil and Eldayem (2014). As these authors further indicate, from 1960 to 1970, mixed-use concepts started to emerge again, activating great urban zones. Later, during 1970–1980, mixed-use development was applied in deteriorated zones to stimulate more liveable areas. Since 1990, mixed-use concepts have been applied more often and are considered a basic element in designing sustainable residential neighbourhoods and applying smart growth principles (Nabil and Eldayem, 2014; Steen, 2016).

According to Urban Hub (2018), the trend in real estate development is more focussed on creating integrated areas by combining different functions instead of developing areas with a homogeneous function. This article further indicates that buildings were previously mainly developed at the micro level, whereas there is now a transition to the creation of a helicopter view and to evaluate an entire area. Such a mixed-use area merges different functions such as residential, retail, commercial, employment, and entertainment functions into a dynamic and comfortable environment in which to stay (MSCR, 2017). A benefit of mixed-use areas is the proximity of functions, which are developed in such a way that people are stimulated to walk to everyday amenities (Moreno et al., 202). This creates a livelier environment since there are more people on the streets during each part of the day, thus stimulating social security (Jacobs, 1961). With more eyes on the street, these areas are less prone to crime and are thus safer (Zahnow, 2018). Ultimately, in a successful mixed-use area, negative externalities vanish and a spill-over of positive externalities arises (Moreno et al., 2020; Zahnow, 2018). In the last decade in particular, an ongoing trend has arisen of people preferring services over products, indicating that the experience component is becoming more essential and should be implemented more in mixed-use areas as well (Alvarez, 2017).

An excellent example indicating the relevance of mixed use on the city level is the ‘15-Minute City’ concept, initially proposed in 2016 by Carlos Moreno, a scientific director and professor in Paris (Moreno et al., 2021). Moreno envisioned a city concept with a high level of mixed use where all residents have access to their basic needs within a 15-minute walk or bicycle ride, stimulating walkable neighbourhoods and communities. According to Moreno, residents will achieve a higher quality of life by having an optimal combination of mixed-use and having the following six essential urban social functions nearby: living, working, commerce, healthcare, education, and entertainment. Due to the COVID-19 pandemic, the 15-Minute City concept has been accelerated considerably according to Moreno. While enforcing strict health protocols and having lockdowns during pandemic, economic activity continued and revealed the vulnerability of cities (Allam and Jones, 2020). Rethinking the city model has become an ever more important subject and cities must remain resilient and liveable in the short and long term whereas the mixed-use concept remains a vital part (Moreno et al., 2021). As the authors further report, Paris will be the first city to fully adopt the 15-Minute City and this concept will be replicated in various cities around the world due to its success, complementing the Smart Cities and the Sustainable Development Goals of the United Nations.

In the Netherlands, the mixed-use concept is well-known and broadly implemented in larger Dutch cities. Mixed use is present at many levels within the city, ranging from mixed-use at the greater level of neighbourhoods to the smaller level of mixed-use buildings with a complete integration of housing and a different set of functions to stimulate interaction within the building (Westbeat, 2020). More importantly, Buck (2020) argued that mixed use should not be seen as a goal on its own, but rather as an instrument to work towards vibrant and future-proof neighbourhoods in cities. From a societal point of view, a relevant topic to investigate is whether mixed use is associated with higher property values and if residents do like the presence of mixed use.

This study uses the Amsterdam market to measure the association between mixed-use and property prices. Up until 2035, Amsterdam is expected to be the fastest growing city in the Netherlands in terms of inhabitants (Planbureau voor de Leefomgeving, 2019). Since mixed use is a common method in (re)developing buildings, neighbourhoods, and cities, applying it will be essential to support the residential growth in Amsterdam (Koomen, Dekkers, and van

Dijk, 2008). Although Amsterdam serves as the case study for this research, the measured association may also indicate how other larger cities in the Netherlands will react to the mixed-use concept. To measure the level of mixed-use in a city, the diversity of employment and housing within neighbourhoods is calculated via a diversity index. Knowledge of the added value of the mixed-use concept is useful for municipalities, real estate developers and investors, it enables these parties to determine which composition of neighbourhoods increases most value to a city. Therefore, from a societal perspective, studying Amsterdam in combination with measuring the additional value for cities by having diverse neighbourhoods is a relevant topic.

1.2 Literature review

The concept of mixed land use became one of the key planning policies in the Western world. In Europe, compact city concepts aim to increase density and mixed land use (Koomen et al., 2008). Earlier studies investigated how mixed use contributes to the value for the area's residents, for employment, or for the surroundings. Nabil and Eldayem (2014) indicated a higher value of living in mixed-use areas since people interact with each other more easily, stimulating social connections and sharing knowledge. Other studies have focused on quantitative analyses and investigated people's willingness to pay to live in mixed-use areas. Song and Knaap (2004) and Koster and Rouwendal (2012) have reported that property prices are valued as higher in mixed-use areas. Cao and Cory (1981) even measured a positive effect in property prices in the surroundings of mixed-use areas. Notably, certain land uses within a mixed-use area (e.g., commercial use, industrial use, education, healthcare, and public parks) contribute differently to property prices. However, when one land use dominates, the area becomes overly monofunctional, being associated with lower property values (Koster and Rouwendal, 2012; Lafferty and Frech, 1978; Song and Knaap, 2004). Therefore, it is essential to select land uses within a mixed-use area carefully for providing a positive value of living.

Within this study, the association between employment and property prices on the neighbourhood level is scientifically relevant since most existing literature has focused on individual property transactions (Cao and Cory, 1982; Lafferty and Frech, 1978; Song and Knaap, 2004; Koster and Rouwendal, 2012). This study investigates how specific employment sectors correlate with residential property prices in the neighbourhood, thereby indicating which employment sectors are found to be most or least valuable to add on the neighbourhood level. Koster and Rouwendal (2012) have focused on the association between property prices and a diverse range of employment in proximity. This study complements Koster and Rouwendal (2012) by investigating whether a diverse range of employment does not only relate to property prices in direct proximity, a neighbourhood for this study, but whether this association also stretches beyond the boundaries of a neighbourhood.[1]

1.3 Problem statement, aim, and research questions

Mixed-use concepts are embedded into European planning policies. Existing literature indicates that the presence of mixed use positively correlates with property prices in the right circumstances. On the other hand, it has not yet been investigated how having a variety of employment sectors, thus the level of mixed use, is associated with property values within a neighbourhood. The aim of this research is to investigate which employment sectors are negatively or positively associated with property values and if more diverse neighbourhoods, in terms of employment sectors, are associated with higher property prices. Thus, the main research question is as follows:

'How does a variety of employment sectors vary with residential property prices within neighbourhoods?'

To support the main research question, the following three subquestions are formulated:

1) According to existing literature, which factors determine property prices in mixed-use areas?

First, to prevent misinterpretations, a mixed-use area is defined. After that, existing literature is consulted to identify which variables determine property prices. In addition to determining different employment and property types, literature is consulted to uncover possible externalities that influence property prices in mixed-use areas.

2) Based on empirical research, which employment sectors correlate positively or negatively within mixed-use neighbourhoods and to what extent?

First, based on empirical research, the diversity of employment and houses on the scale of neighbourhoods is identified for Amsterdam by using a diversity index. Subsequently, the presence of different employment sectors within Amsterdam's neighbourhoods is defined. Finally, to what extent these employment sectors correlate positively or negatively with the average property prices within the neighbourhoods is investigated.

[1] No existing literature has been found by using the searching machines SmartCat (the library catalogue of the Library of the University of Groningen) and Google Scholar, by using the terms: 'Property valuation mixed use', 'Property valuation mixed use neighbourhoods', 'Valuation mixed use neighbourhoods', 'Mixed use property prices', 'Mixed use development property prices Amsterdam', 'Diversity index neighbourhoods', 'Diversity index property prices', and 'Diversity index property valuation neighbourhoods'.

3) To what extent is a greater variety of employment sectors in surrounding neighbourhoods associated with property prices?

To determine whether neighbourhoods are associated with a wider variety of employment sectors, a diversity index is applied. In addition, it is analysed whether having a wider variety of employment sectors in surrounding neighbourhoods is also associated with property prices. Doing so might empirically indicate that property prices are determined by characteristics that reach beyond the boundaries of a neighbourhood itself.

1.4 Conceptual model

The conceptual model in Figure 1 visualises the association between the variables. The main question is whether different employment sectors vary with the value of residential properties in neighbourhoods. This conceptual model visualises the association between the independent variable (i.e., employment sectors in mixed-use neighbourhoods and the diversity index) and the dependent variable (i.e., property price). To increase confidence in the outcome of this research, control variables are applied (i.e., structural, spatial, and social environmental characteristics).

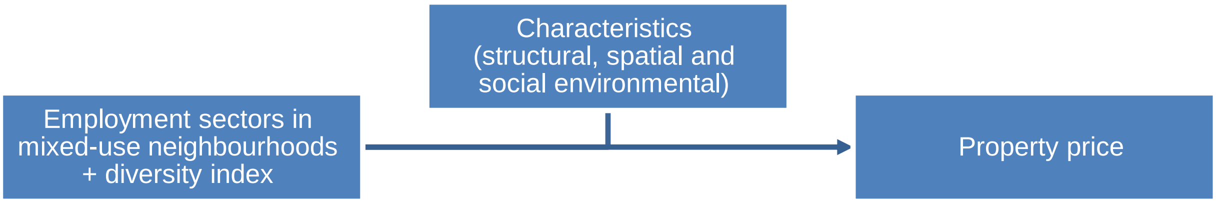


Figure 1: Conceptual model

1.5 Methodology and data

Within this study, a cross-sectional approach is conducted, combining three datasets. First, a dataset from Statistics Netherlands (CBS) is used, including a wide variety of information of the Amsterdam neighbourhoods. Second, the LISA dataset is used, containing information of the size and type of employment for all businesses in Amsterdam. The third dataset that is used originates from the Key register Addresses and Buildings (in Dutch: Basisregistratie Adressen en Gebouwen, BAG), indicating structural property characteristics. By using ArcMap GIS, these datasets are combined based on the geographical location, resulting in a multiple linear regression model.

1.6 Thesis overview

The remainder of this thesis is organised as follows. Chapter 2 (Theoretical Framework) presents the theory that this study was grounded in, reflects upon what the relevant literature has achieved to date, and proposes the study's hypotheses. Chapter 3 (Data and Method) describes how the linear models were constructed before discussing the operationalisation given the available data. Chapter 4 (Results) presents the results of the linear regression models and interprets them; furthermore, it describes the sensitivity analysis that was performed to determine whether the dependent variable would perform differently if the independent variables underwent changes. Finally, Chapter 5 (Conclusion and Discussion) finalises the thesis and provides recommendations for future research.

2. THEORETICAL FRAMEWORK

This chapter presents the theoretical framework, which consists of three topics. These three topics are presented in the first three sections of this chapter. First, Section 2.1 elaborates the mixed-use concept, providing a definition, explaining its origin, and discussing how mixed-use applies to practice. Second, Section 2.2 focuses on property values, which determine characteristics and externalities of property values which hold for mixed-use areas as well. Third, Section 2.3 concerns employment sectors, whereas the effect of employment sectors on property values is analysed in mixed-use areas. Finally, the Section 2.4 presents the hypotheses.

2.1 Mixed-use in theory

Definition

The main goal of applying the mixed-use concept is to positively contribute to an area's direct and indirect surroundings by combining different functions. Coupland (1997) defined a mixed-use area as an area in which the concentration and activities promote vitality, bustling town centres exist, cars are less relied on, and travel time is reduced. Vorontsova et al. (2016) defined mixed use in terms of functions, namely as a combination of residential, commercial, institutional, cultural, and production functions, which allows people to live, work, relax, and shop in one place. All of these factors help to improve environmental quality, social equity, and economic strength (Grant, 2002). The area is vibrant every day of the week and during each part of the day due to the different functions therein, thus ensuring more people on the streets (Jacobs, 1961). The area contributes to the sustainable city concept, but also creates more eyes on the street, which enhances social control and public safety (Jabareen, 2006; Jacobs, 1961; Zahnow, 2018). In the present study, mixed use is defined as 'neighbourhoods with different employment sectors and housing'. The vibrancy of the neighbourhood, higher public safety, and many amenities in proximity to the housing enhance the value of the entire neighbourhood and its users. This research specifically tests whether this varies with property values.

Origin of the mixed-use concept

Jane Jacobs, a well-known early critic, noticed problems in the urban planning policies of the 1950s. Entire building blocks were cleared and rebuilt, ignoring everything that made cities great, and many neighbourhoods in the United States declined (Jacobs, 1961). She performed a pioneering analysis on what contributes to successful neighbourhoods. In *The Death and Life of American Cities* (1961), she encouraged the integration of diversity and mixed-use buildings, arguing that a successful and vibrant neighbourhood is formed from the diverse use of fine-grain combinations. According to Jacobs (1961), mixed use is about finding a balance between work, service, and living activities that provides a stimulating, lively, and secure public realm. In the years after her publication, the negative impact of urban renewal became more evident and the ideas of Jacobs seemed increasingly cogent (Grant, 2002).

Just over a decade after Jacobs' publication, the concept of the compact city was introduced by Dantzig and Saaty (1973), which contained the basic ideas introduced by Jacobs. They envisioned a sustainable city model that enhanced quality of life but not at the cost of the next generation. The compactness encompassed density in the built environment and the intensification of its activities, efficient planning, diverse and mixed land uses, and efficient transportation systems. In social terms, compactness and mixed uses are associated with diversity, social cohesion, and cultural development (Jabareen, 2006). In 1990, the European Commissions' Green Paper strongly advocated the compact city concept, assuming an improved quality of life and more environmentally sustainable urban areas (Commission of European Communities, 1990). In a compact city, the European Commission proposed reduced travel distances, saving of rural land from development, support for local facilities, and the creation of more autonomous local areas. Partly due to the compact city model, the mixed-use concept became widely accepted and is one of the key planning policies in the Western world (Jabareen, 2006; Koomen et al., 2008; Koster and Rouwendal, 2012; Rowley, 1996).

Mixed-use concept in practice

Mixed use is present at different levels within cities. Jacobs (1961) defined mixed use at the neighbourhood level, whereas Grant (2002) focused on the local level and Coupland (1997) at the level of the building complex itself. For clarification, Rowley (1996) defined four settings in which mixed use occurs: (1) within districts or neighbourhoods, (2) within the street and other public places, (3) within the street or building block, and (4) within individual buildings. In addition to the level of mixed use, Hoppenbrouwer and Louw (2005) defined different forms of mixed use, reporting four dimensions. The first dimension was shared premises, which combines two functions within the same premises, such as self-employed people. The second was the horizontal dimension, in which one building is intended for work and the building directly next to it for housing. Third was the vertical dimension, which has for example commercial uses on the ground floor with multiple layers of housing on top. The fourth and final dimension was time, where a particular space is intended for different functions on different days or parts of the day. For example, certain spaces within a theatre could be used for conferences during the daytime and serve as a cinema in the evening.

To advocate for mixed-use development, different objectives and strategies are used, for which Grant (2002) formulated three conceptual levels. The first is to increase the intensity of land uses in terms of housing types to stimulate a social mix; the second is to increase diversity of uses to stimulate synergies and to avoid conflicts between uses; and the third is to integrate segregated uses by bringing categories of use together to overcome regulatory barriers. Using these conceptual levels as guidance are helpful since implementing mixed use is not always self-evident as reported by Nozeman (1977). Due to urban renewal, not much space is left for spontaneous developments or mixing functions. His study focussed on retail facilities and influential parties such as urban planners and real estate developers prefer clustered over scattered facilities. Clustered retail facilities result in relatively lower construction costs due to their greater scale, which leads to higher rent prices for retailers and easier acquisition of land (Nozeman, 1977). Therefore, to stimulate a livelier environment by mixing functions involves effective coordination between parties at various scales.

To generate ‘exuberant diversity’ in a city’s streets and districts, Jacobs (1961) listed four well-known indispensable conditions. First, the district must serve more than one primary function, and preferably more than two; second, building blocks must be short and contain frequent turns; third, the district must mingle close-grained buildings that vary in age and condition; and fourth, there must be a sufficiently dense concentration of people. When mixed use is successfully implemented, it increases the value of living (Jacobs, 1961). Therefore, combining multiple functions encourages economic growth due to the interaction between different industries (Liusman et al., 2017). For example, Nabil and Eldayem (2014) indicated that a strong relationship exists between the number of land uses and social capital in the city of Cairo, Egypt. Social capital is defined as the working product of interpersonal networks, contacts, and related human resources. This contributes to stimulating more interaction and creates a spirit of teamwork and cooperation between the users of the area, leading to a higher value of living.

2.2 Values of properties

Elements for constructing property prices in the presence of mixed use

Many publications have been written about impact or effect of mixed use from a qualitative point of view (Coupland, 1997; Dantzig and Saaty, 1973; Grant, 2002; Hoppenbrouwer and Louw, 2005; Jabareen; 2006; Jacobs, 1961; Nozeman, 1977; Rowley, 1996). Less common in the literature is the application of quantitative methods, with only a few studies actually measuring the effect of mixed use on property prices (Cao and Cory, 1982; Koster and Rouwendal, 2012; Lafferty and Frech, 1978; Song and Knaap, 2004). To understand which variables are useful to apply in this research, Appendix 1 indicates which variables the authors of quantitative research studies have used to measure the effect of property prices in the presence of mixed use.

External effects on property prices

To determine property prices, external effects are essential to consider since they may play a significant role. The focus of these externalities relates to buildings in general and therefore might explain potential variations in the property prices in mixed-use neighbourhoods as well. Externalities could be negative as well as positive. In terms of negative externalities, Lafferty and Frech (1978) measured the strength of externalities and indicated negative ones when buildings or land are vacant as well as if land uses are scattered. Palmquist (1992) found negative externalities for living close to highways due to noise. Moreover, an area with a high density of households would not be preferred since it leads to higher crime rates and reduced privacy (Glaeser et al., 2005).

Positive externalities exist when multifamily land use is suitably concentrated, resulting in higher property values (Lafferty and Frech, 1978). Other studies have suggested that positive externalities are also derived from the presence of open and public accessible spaces such as parks, which are also observed to increase property values (Jang and Kang, 2015; Koster and Rouwendal, 2012; Song and Knaap, 2004). Daams et al. (2016) reported a premium on property prices if housing in the Netherlands is located close to attractive natural spaces, e.g. forests, water spaces, parks, and recreation areas. In addition, Fruth et al. (2019) indicated that urban green space supports physical and mental health due to recreation opportunities, aesthetics, and the regulation of air quality levels. The presence of water is also appreciated by households according to Rouwendal et al. (2014); households located directly next to water pay a premium of 5% for their property, with the effect being measurable up to 60 metres.

Externalities may or may not be appreciated by residents and potentially be associated with higher or lower property values. When analysing property prices and having available data on the aforementioned externalities might provide helpful insights. Based on existing literature, the next section discusses specifically how mixed use is related to property prices.

2.3 Mixed use and property prices

Influence of mixed use on property prices

The value of living could also be indicated in terms of property value or people's willingness to pay. Cao and Cory (1981) researched whether an effect in property prices is measurable in surrounding properties of mixed-use areas in Tucson, Arizona. They measured a positive effect on surrounding properties, indicating that the reach of mixed use even goes beyond the boundaries of the mixed-use area itself. Song and Knaap (2004) researched the impact of mixed use on the property prices of single-family homes in Washington county. They found an increased value in single-family homes in areas where non-residential land uses are evenly distributed and where more service jobs are available. From another point of view, Koster and Rouwendal (2012) conducted research in the Rotterdam market and analysed whether property values increase when the property is located in mixed-use areas. They reported increased property values with the presence of mixed-use.

In addition, residential property types contribute differently within mixed-use areas. By adding multifamily homes within a mixed-use area, the value of surrounding properties tends to increase more than when single-family homes are added, as indicated by Cao and Cory (1981). This aligns with Song and Knaap (2004) and Koster and Rouwendal (2012), who have reported that single-family homes are valued lower when multifamily homes are nearby. The reason is that single-family homeowners appreciate homogeneous residential neighbourhoods; for example, they prefer living in a low address density environment with larger plots of land. By contrast, multifamily homeowners are willing to pay more for living in a more diversified area in terms of amenities, which makes sense since most apartments are located in urban areas with a higher address density, thus having many amenities close by (Cao and Cory, 1981; Koster and Rouwendal, 2012; Song and Knaap, 2004).

Effect of employment sectors on property prices in mixed-use areas

The proximity of some types of land use contribute substantially more to people's willingness to pay in mixed-use areas, starting with the employment sectors that are valued as positive (Cao and Cory, 1981). As these authors indicated, the value of surrounding properties increases when the share of industrial, commercial, multifamily, and public land-use activity in a neighbourhood increases. Song and Knaap (2004) also found a positive relationship between property prices and commercial use nearby, supporting the 15-Minute City concept mentioned in the introduction. These categories are somewhat more abstract compared with those of Koster and Rouwendal (2012), who specified more types of land use. They reported that the employment sectors of business services, education, healthcare, leisure, and retail are valued positively by households. From another point of view, having a greater mix of functions close by also translates into significantly higher office rents (Liusman et al., 2017).

According to Koster and Rouwendal (2012), employment sectors that contribute negatively to property prices are schools in the direct vicinity as well as manufacturing, government, and wholesale, which is in line with the findings of Verwolf (2019). An additional employment sector that negatively impacts property prices is the agricultural sector (Laferty and Frech, 1978). In contrast to the findings of Cao and Cory (1981), De Vor and De Groot (2011) measured negative effects in residential property prices when industrial employment sectors are close by. Jang and Kang (2015) reported a negative effect on house prices in the proximity of hypermarkets.[2] . For each of the individual employment sectors, a similar principle applies; when the area becomes too monofunctional and one employment sector dominates, it is associated with lower property values (Koster and Rouwendal, 2012; Lafferty and Frech, 1978; Song and Knaap, 2004).

From another point of view, Verwolf (2019) analysed vacancy in retail areas and investigated which functions provide the highest contribution by filling vacancies and future-proofing retail areas. The author measured this by not only a financial component, but covered the synergy between functions, physical integration, a legal component and market conditions as well. The most suitable function to add to the retail area was found to be catering, followed by housing, fitness, offices, primary care, cinema, and finally day-care. The following functions were considered unsuitable to add to retail areas: education, government, wholesale, and industry. Since Verwolf (2019) combined different functions in one area, including housing, this already provides an indication of which functions are also suitable in mixed-use areas.

The successful establishment of a mixed-use area in which property prices increase depends on numerous factors. An optimal combination of different types of land use must be sought since each area has its own characteristics and specifications. The amount of economic activity must be scaled in size to fit well within the area (Cao and Cory, 1981; Song and Knaap, 2004). Although Nozeman (1977) focussed on retail facilities, this accords with his findings since he reported that there is no optimal distribution pattern of facilities within an area, and each area has to be assessed individually. An optimal combination of mixed use increases the property prices with 6% compared to monofunctional residential areas (Koster and Rouwendal, 2012).

2.4 Hypotheses

Based on the aforementioned literature, this study proposes the following hypotheses:

Hypothesis 1 (H_1): There is a positive association between average property prices and the presence of commercial, education, healthcare, and leisure employment sectors.

---

[2] Large scale supermarket where you can buy groceries as well as department store items.

This hypothesis is supported by the studies of Cao and Cory (1981), Koster and Rouwendal (2012), and Song and Knaap (2004). These authors all found that the presence of the commercial, education, healthcare, and leisure employment sectors correlate positively to property prices in mixed-use areas.

Hypothesis 2 (H_2): There is a negative association between average property prices and the presence of agriculture, industry, wholesale, and government employment sectors.

The employment sectors of agriculture, industry, wholesale, and government are negatively associated with property prices in mixed-use areas (Koster and Rouwendal, 2012; Lafferty and Frech, 1978; Verwolf, 2019). The association in the industrial sector are ambiguous because positive (Cao and Cory, 1981) and negative (De Vor and De Groot, 2011; Koster and Rouwendal, 2012) associations on property prices have been found; however, the research of De Vor and De Groot (2011) and Koster and Rouwendal (2012) focused on the Netherlands. Therefore, this study expects that industrial employment sectors are negatively associated with property prices in mixed-use areas.

Hypothesis 3 (H_3): The association between the average property prices and the presence of employment is stronger for multifamily homes compared to single-family homes.

That a higher share of multifamily homes is stronger associated with higher property values in mixed-use areas compared with a higher share of single-family homes is supported by the studies of Cao and Cory, (1981), Koster and Rouwendal (2012), and Song and Knaap (2004).

Hypothesis 4 (H_4): There is a positive association between average property prices and a greater diversity of employment sectors.

Having a greater range of diversity in terms of employment sectors in combination with housing achieves the highest value of living, as expressed in property values in mixed-use areas, compared with a dominant employment sector in comparable areas (Cao and Cory, 1981; Koster and Rouwendal, 2012; Lafferty and Frech, 1978; Nozeman, 1977; Song and Knaap, 2004).

3. DATA AND METHOD

This chapter focusses on the data and the conducted method of this study. First, Section 3.1 focusses on the data and indicates how this study relates to existing literature. Section 3.2 elaborates on the procedure of calculating a diversity index for a neighbourhood, thereby indicating the level of mixed-use in a neighbourhood, and the calculation of a diversity index of surrounding neighbourhoods. The linear regression models are the focus of Section 3.3, describing the ordinary least squared assumptions and the established regression models. Section 3.4 indicates the descriptive statistics of this research. To finalize, Section 3.5 discusses the essence of performing a sensitivity analysis.

In this study, a cross-sectional approach is employed to measure the association of the independent variables of employment sectors and the diversity index on the dependent variable of price. The real estate market does not change overnight, indicating that a trend over multiple years is not likely to occur and observing one year is more appropriate (Brooks and Tsolacos, 2010). The year 2018 is used for this research since the most recent version of one of the datasets is from 2018. Relevant studies have deemed the most suitable methodology for investigating effects on property prices within mixed-use neighbourhoods to be a linear regression model (Cao and Cory, 1982; Laferty and Frech, 1978; Koster and Rouwendal, 2012; Song and Knaap, 2004).

3.1 Data

<u>Variables</u>

This study uses three datasets, a dataset from the Statistics Netherlands (CBS), the LISA dataset, and the Key register Addresses and Buildings (in Dutch: Basisregistratie Adressen en Gebouwen, BAG). The dataset of the CBS contains a wide variety of information of the neighbourhoods in Amsterdam, including the boundaries of all 481 neighbourhoods. The dependent variable for this research originates from the CBS dataset, namely the average value of immovable property per neighbourhood, or the WOZ value. Municipalities determine this value by means of appraisal, where building characteristics and the location of buildings are compared to one another (Rijksoverheid, 2021). The WOZ value is a well-accepted estimate for property prices in the Netherlands and is also comprehensive since the average value is calculated for owner-occupied homes as well as rental homes. The dataset does not have an average WOZ value for neighbourhoods when the neighbourhood has fewer than 20 residential properties, or fewer than 50 WOZ objects.[3] These restrictions result in 382

[3] A WOZ object is defined as a residential property that includes room for work-related services (CBS, 2020).

remaining neighbourhoods that contain on average 1125 homes and 2200 residents. Appendix 2a visualises all neighbourhoods in Amsterdam and indicates the average WOZ value per neighbourhood. Besides the WOZ value, the CBS dataset contains many other variables on the neighbourhood level, including variables relating to the variables in existing literature as described in Appendix 1.

The LISA dataset defines the employment sectors according to the Dutch SBI (Standard Industrial Classification) standards, which is a hierarchical classification of 21 economic activities. The CBS merged these activities according to the employment sector, reducing it to eight categories. These reduced categories are used in this study since it would be easier to comprehend than 21 categories. These data contain the number of employees per sector per neighbourhood.

The BAG dataset consists of information of individual properties. The floor space and construction year per property are used from this dataset and consequently divided into cohorts. By doing this, each cohort group contains percentages and enables easier comparison between the different cohort groups. This indicates for example the share of properties per neighbourhood that belongs to a floor area between 76 and 100 square metre, or construction period between 1900 and 1944. By using the program ArcMap GIS, the CBS, LISA, and BAG dataset are merged per neighbourhood, based on the geographical location.

The three datasets do not cover all of the variables from earlier literature presented in Appendix 1. On the other hand, the CBS dataset does contain some other potentially relevant variables. All of the used variables are presented in Appendix 3. Based on the three datasets, the independent and control variables are divided into the following categories:

- Employment sectors in the neighbourhood;
- Housing characteristics in the neighbourhood;
- Spatial characteristics of the neighbourhood;
- Social environment of the neighbourhood.

The employment sectors are completely covered by using the LISA dataset as described earlier in this section. For housing characteristics, the number of homes, floor space, construction year, and property type overlapped with those in earlier literature. The variable of floor space serves as proxy for lot size and volume. This research is based on neighbourhoods with average values, whereas earlier literature has used individual observations of properties. Some of the variables are not precisely translatable; therefore, no variables are available for rooms, garages, basements, gardens, central heating, or listed buildings. Whereas Koster and

Rouwendal (2012) specified four types of homes, the present study only has data available for a distinction between multi- and single-family homes. In addition, this study contributes to literature to add the variables of inhabited housing, vacant housing, owner-occupied housing, and rental housing, and checks whether these variables are also associated with average property prices.

In terms of spatial characteristics, only the dataset of the CBS is conducted. The variables of distance to train station and distance to main road are used, similar as in the study of Song and Knaap (2004) and Koster and Rouwendal (2012). For distance to the centre, to the CBD (central business district), to minor roads and to open spaces, the CBS dataset does not have data available, and therefore, they are not covered in this research. Song and Knaap (2004) mentioned the variables of proximity to water bodies and proximity to bus stops. This research measures the presence of water differently by examining the share of water within the neighbourhood. Data for bus stops are not available, but data including distance to important transfer stations are available and therefore used. In addition, data are available for the average distance to 29 different amenities. To prevent omitted variable bias, all amenities are included in the analysis as presented in Appendix 3, under the heading ‘Spatial characteristics in the neighbourhood’.

Regarding the social environment of the neighbourhood, the variables of population density, multifamily households, ethnic minorities, low-income households, employment status, and crime rates are in line with earlier literature. For the variables of noise and crowdedness, no data are available and is therefore not considered. The variable of income was present in the CBS dataset despite it contained many missing variables, which led to this variable being excluded from the present research. The following extra variables are also added to this research: average address density per square kilometre, urbanity level, high-income households, households below the social minimum, three types of social security benefit receivers, and three types of crime records. A summary of all variables is presented in the table in Appendix 3.

3.2 Diversity index

Diversity within the neighbourhood

To determine whether a mixed-use neighbourhood is diverse in terms of employment sectors, a diversity index is used. The inverse of the Hirschman–Herfindahl index was used by Koster and Rouwendal (2012), whereas the Herfindahl index was used by Wo and Kim (2020). These indices are similar since both measure the proportion of different employment sectors within the neighbourhood. The Hirschman–Herfindahl index is more advanced since it also considers the proportion of households within the area and measures whether households dominate the neighbourhood as well. When the term 'diversity index' is used in this thesis, it refers to the inverse of the Hirschman–Herfindahl index, the formula for which is as follows:

$$D_n = \frac{1}{\sum_{\forall g}\left(P_{gn}^2\right) + P_{hn}^2} \tag{1}$$

Where D_n is the diversity index for neighbourhood n; P_{gh} is the employment share of the sum of employment; and P_{Hh} is the household share of the sum of households. These are defined in the following formulas:

$$P_{gn} = \frac{E_{gn}}{H_n + \sum_{\forall g}(E_{gn})} \tag{2}$$

$$P_{hn} = \frac{H_n}{H_n + \sum_{\forall g}(E_{gn})} \tag{3}$$

Here, the variable E_{gn} accounts for the number of employees in sector g and neighbourhood n, whereas H_n is the number of households in neighbourhood n. There is no diversity when the diversity index has an outcome of 1 and the neighbourhood is completely dominated by either one employment sector or households. The higher the outcome of the diversity index, the higher the diversity of employment sectors and households within the neighbourhood. The diversity index has a mean of 2.69 and a standard deviation of 1.02.

In addition, it may appear surprising that the number of employees is used in combination with the number of households instead of the number of residents. Nevertheless, comparing these has been a common measure in earlier research (Cao and Cory, 1982; Koster and Rouwendal, 2012). No exact reason has been given, but the most likely reason is that households are directly related to the level of house prices, making a comparison of the number of employees with the number of households relevant.

<u>Diversity of surrounding neighbourhoods</u>

A neighbourhood does not solely depend on the characteristics within it – it may also benefit from the characteristics of surrounding neighbourhoods. To account for this, the model is extended using a spatial variable that not only includes the characteristics of the 'own' neighbourhood but also incorporates the weighted sum of characteristics in the directly surrounding neighbourhoods (Van Duijn and Rouwendal, 2013). The surrounding neighbourhoods only consist of neighbourhoods that have residents in their neighbourhood as well, which equals neighbourhoods *n* as used in the later models. By including surrounding neighbourhoods without housing, the outcome of the diversity index is disproportionally high and upwardly biased since it calculates diversity within the neighbourhood among employment sectors and does not account for housing. The formula of this spatial variable is as follows:

$$Da_n = \left(\sum\nolimits_{m \epsilon C_n} D_m\right) / m \tag{4}$$

The variable Da_n is the average diversity index of the neighbourhoods surrounding neighbourhood *n*. The sum of the diversity index D_m is taken, whereas *m* stands for the surrounding neighbourhoods of neighbourhood *n*, and C_n stands for the specified set of neighbourhoods surrounding neighbourhood *n*.

The variable Da_n calculates the average of the diversity index of the surrounding neighbourhoods, indicating that it sums the outcomes of the diversity indices and subsequently divides it by the number of surrounding neighbourhoods. Some neighbourhoods have more employees than others and it is expected that more activity occurs between such neighbourhoods. To account for this, a new variable is created that puts a weight on the surrounding neighbourhoods, whereas the weight is based on the number of employees in the surrounding neighbourhoods. This provides a weighted average of the diversity index for each neighbourhood, resulting in the following formula:

$$Dwa_n = \sum\nolimits_{m \epsilon C_n} \left(\frac{E_m}{\sum_{C_n} E_m}\right) \; D_m \tag{5}$$

The variable Dwa_n is the weighted average of the diversity index of the surrounding neighbourhoods of neighbourhood *n*. Furthermore, E_m is the number of employees in neighbourhood *m*; *m* stands for the surrounding neighbourhoods of neighbourhood *n*; and C_n stands for the specified set of neighbourhoods surrounding neighbourhood *n*. The weighted average is calculated by taking the share of employment, or E_m, in neighbourhood *m* and dividing that by the sum of employment of the neighbourhoods surrounding neighbourhood *n* in the specified set C_n. Finally, the weight is multiplied by the diversity index of the surrounding neighbourhood *m*.

This study also considered to create another variant of the variable Dwa_n, providing a weight on the average diversity index based on the length of common borderlines of surrounding neighbourhoods. In this situation, the formula is similar as in equation 5, where E_m is replaced by the length of the border line. Large common borderlines give those surrounding neighbourhoods a larger weight in the average diversity index compared with neighbourhoods that only have a common borderline of several metres. The weight of the number of employees per neighbourhood is chosen over the common borderline and, therefore, this variant is not used as variable. This research focuses on the association between residential property prices and employment, but also on the diversity of employment. Therefore, it is expected that the number of employees more accurately represent and characterise a neighbourhood compared with measuring the pure geographical location.

3.3 Linear regression models

<u>Assumptions in linear regressions</u>

In earlier literature, hedonic modelling has often been used. This research is based on average values of neighbourhoods, and therefore, it uses aggregated data. Therefore, the estimation technique adopted for this research is a linear regression model. More specifically, this study concerns a multiple linear regression model since describing and evaluating the relationship between two or more independent variables and one dependent variable (Brooks and Tsolacos, 2010). The dependent variable in this research is the average WOZ value of properties per neighbourhood and is constructed based on many different aspects, particularly its physical characteristics, location, and social environment, making it vital to control for these aspects as well. The independent variables are the diversity index, present employment sectors in the neighbourhood, its physical characteristics, average distance to amenities for residents, and its social environment. Eventually, the average WOZ value of properties is linked to the diversity of employment sectors in the neighbourhood, attempting to explain the marginal price that residents are willing to pay for living in diverse neighbourhoods, whereas the marginal price is based on the aggregated data of the neighbourhoods. According to Brooks and Tsolacos (2010), the most common method for fitting a line to the data is ordinary least squares (OLS). To perform OLS, the authors indicated that the assumptions presented in Table 1 must hold concerning the error term.

Within this research, some model adjustments are made by excluding variables and interpreting some as dummy variables. In addition, a critical factor is the absence of multicollinearity. Appendix 4 elaborates on these factors further. Subsequently, all models are tested for the OLS assumptions. Appendix 5 represents the performed tests, indicates how

Table 1 : OLS assumptions (Brooks and Tsolacos, 2010).

Technical assumption	***Interpretation***	***Referring to***
1. $E(u_t) = 0$	The errors have zero mean.	Linearity
2. $var(u_t) = \sigma^2 < \infty$	The variance of the errors is constant and finite over all values of xt.	Homoscedasticity
3. $cov(u_i, u_j) = 0$	The errors are statistically independent of one another.	No autocorrelation
4. $cov(u_t, x_t) = 0$	There is no relationship between the error and corresponding x variable.	Non-stochastic and exogeneity
5. $u_t \sim N(0, \sigma^2)$	ut is normally distributed.	Normality

these assumptions are tested and what procedures are taken for each of the OLS assumptions. Assumptions 1, 2, 3 and 5 hold, but assumption 4 is violated due to the presence of endogeneity as indicated in performed test. According to Brooks and Tsolacos (2010), the coefficients are therefore not consistent and the standard errors not efficient, indicating that the coefficient estimates are no longer BLUE (best linear unbiased estimator). Due to this violation, this research does not serve as an impact study and the effects or causal relationships between variables cannot be interpreted. Consequently, only the association and the strength of the association between variables can be interpreted.

Models

The final linear regression model is constructed by using different models. Note that this study uses the program Stata to perform the statistical regression models, the output of the entire syntax is presented in the final Appendix 15. The first model, the baseline model, contains the most basic characteristics, whereas every successive model contains new variables or a variation in the used variables. In every model, this study checks how well the regression models actually fit the data by examining the explained variance, or R^2. It is beneficial when the R^2 increases since it implies that more of the variation in the dependent variable can be explained by the variation in the independent variables (Brooks and Tsolacos, 2010). Furthermore, all models include the intercept α, the coefficients β to be estimated, and the error term ε_n. The average WOZ-value per neighbourhood is checked for normality by using histograms. It appeared that using the exact values for average property values per neighbourhood is not normally distributed, therefore this variable is log transformed, Appendix 6 represents the log transformation. The first model is constructed with the dependent variable

$\ln P_n$, which denotes the logarithm of the average WOZ value for properties in neighbourhood *n* and the independent variable D_n, which is the diversity index of neighbourhood *n*, resulting in the following baseline model:

$$\ln P_n = \alpha + \beta_1 D_n + \varepsilon_n \tag{1}$$

The second model also includes the average value of the diversity index of surrounding neighbourhoods and is divided in two versions. The first version includes the variable of the average diversity index of surrounding neighbourhoods as mentioned in equation 4, or Da_n.The second version includes another variant of this variable, the weighted average of the diversity index of surrounding neighbourhoods as formulated in equation 5, or Dwa_n. This results in the following models:

$$\ln P_n = \alpha + \beta_1 D_n + \beta_2 Da_n + \varepsilon_n \tag{2a}$$
$$\ln P_n = \alpha + \beta_1 D_n + \beta_2 Dwa_n + \varepsilon_n \tag{2b}$$

As explained in Section 3.2, the variable Dwa_n was expected to more accurately represent the average value of the diversity index since it connects a weight to the number of employees in the surrounding neighbourhoods. It appears that this variable does not result in more explained variance.[4] Therefore, the average diversity index, Da_n, is considered more suitable than the weighted average of the diversity index, Dwa_n. For this reason, Da_n is added to the baseline model. The third model is expanded by including vector X_n for the independent variables of different employment sectors in neighbourhood *n*, to provide insights in how the employment sectors are associated with the average WOZ value, Within the calculation of the diversity index, a calculation is also performed using the employment sectors. This can be assumed not to cause implications between the variables since Koster and Rouwendal (2012) used the same method and did not encounter any problems. The resulting model is as follows:

$$\ln P_n = \alpha + \beta_1 D_n + \beta_2 Dwa_n + \beta_3 X_n + \varepsilon_n \tag{3}$$

[4] It is tested whether the average diversity index, $\mathrm{Da_n}$, or the weighted average of the diversity index, $\mathrm{Dwa_n}$, is more representable evaluating the explained variance, or R-squared. Regression model 2a and 2b are represented in Table 3. More activity was expected between neighbourhoods with more employees. Consequently, it was expected that this would lead to a stronger association between the diversity index and neighbourhoods with more employees. It appeared that the R^2 of model 2a is higher in comparison with model 2b, indicating that more explained variance of the dependent variable of the average WOZ value in the neighbourhood. Because of this, model 2a is more likely to be closer to the true value of the coefficient. This contradicts the expectation that more employment in surrounding neighbourhoods better represents activity from and to these neighbourhoods. For this reason, the average diversity index (or avg_di) is chosen for each successive model instead of the weighted diversity index according to the number of employees (or w_di_ba).

The fourth model includes vector Y_n, which serves as a control variable for housing characteristics in neighbourhood *n* to provide a better explanation of the average WOZ value in the neighbourhood. This results in the following model:

$$\ln P_n = \alpha + \beta_1 D_n + \beta_2 Dwa_n + \beta_3 X_n + \beta_4 Y_n + \varepsilon_n \quad (4)$$

The fifth model also accounts for the control variables of spatial characteristics in neighbourhood *n* by adding vector Z_n. As indicated in the literature, proximity to transportation options may affect property prices. To control for the association of proximity to amenities, the average distance to amenities is covered in vector Z_n, resulting in the following model:

$$\ln P_n = \alpha + \beta_1 D_n + \beta_2 Dwa_n + \beta_3 X_n + \beta_4 Y_n + \beta_5 Z_n + \varepsilon_n \quad (5)$$

The sixth model includes control variables in the social environment of neighbourhood *n* by extending the formula with vector μ_n. The social environment includes the variables of residential characteristics, income-related characteristics of residents, and criminality in the neighbourhood, resulting in the following model:

$$\ln P_n = \alpha + \beta_1 D_n + \beta_2 Dwa_n + \beta_3 X_n + \beta_4 Y_n + \beta_5 Z_n + \beta_5 \mu_n + \varepsilon_n \quad (6)$$

3.4 Descriptive statistics

Appendix 7 presents the descriptive statistics of all variables used in this research, Table 2 summarizes the most relevant descriptive statistics. Important to note is that the distance to amenities is measured on paved roads for cars, not on bicycle or pedestrian paths (CBS, 2020). Since Amsterdam is highly suitable for walking and cycling and less suitable for cars, the distance variables might seem somewhat high. For example, the average distance to a supermarket (*af_superm*) is more than 500 metres, which is quite high considering the many supermarkets in Amsterdam and the high urban level. Another noteworthy phenomenon is that just over half of the residents in Amsterdam has a migration background (variable *p_west_al* and *p_n_w_a*). This is relatively high compared with the Dutch average, where almost a quarter of the population has a migration background (CBS, 2019). In addition, almost half of the households in Amsterdam belongs to the lowest income group (*p_laaginkh*); this variable is calculated for people who belong to the lowest 40%-income group in the Netherlands. Amsterdam has 10% more residents in the lowest income group compared with the Dutch average. Moreover, 18% of households in Amsterdam belong to the highest income group

Table 2 : Summary of the most relevant descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
log of average woz value	382	12.709	.468	11.067	14.423
diversity index	382	2.691	1.024	1.212	6.383
average diversity index	382	2.719	.674	1.422	4.550
weighted average diversity index	382	2.922	.753	1.431	5.106
employees sector a (x100)	382	.005	.025	0	.31
employees sector bcdef (x100)	382	.524	1.25	0	16.45
employees sector gi (x100)	382	2.849	3.895	0	30.71
employees sector hj (x100)	382	1.11	1.582	0	13.09
employees sector kl (x100)	382	1.052	6.437	0	96.78
employees sector mn (x100)	382	2.692	4.007	.02	48.26
employees sector opq (x100)	382	2.758	4.524	.01	53.66
employees sector rstu (x100)	382	1.161	1.137	.01	9.92
homes	382	1123.937	668.556	53	3359

(*p_hooginkh*), whereas this variable is calculated for the highest 20%-income group in the Netherlands. A complete elaboration of all variable names is provided in Appendix 3.

3.5 Sensitivity analysis

To test whether a different setup of the independent variables would still lead to the same results, a sensitivity analysis is applied. Within this analysis, four different robustness tests are conducted. The first test checks whether the number of businesses leads to other results compared with the number of employees. The second test uses all 21 employment sectors as single entities instead of the eight employment sectors currently merged as mentioned in Section 3.1. A complete description of the 21 employment sectors is provided in Appendix 3. This robustness test aims at providing an enhanced understanding of which specific types of employment correlate with the average property price in a neighbourhood. The third robustness test checks for a nonlinear association between the employment sectors and the average property price per neighbourhood by squaring the data of the employment sectors. The final test is a Chow test that compares the subgroups with a pooled regression model, each subgroup represents a group with a different population density.

4. RESULTS

This chapter presents the results of the study and is structured according to the hypotheses described in Section 2.4. First, Sections 4.1 to 4.4 examine the four hypotheses. In addition to the variables described in the hypotheses, other variables also had coefficients significantly different from zero, which are elaborated further in Section 4.5. The final part of the chapter – Sections 4.6 to 4.9 – focuses on the four robustness tests described in Section 3.5.

Before starting to elaborate on the hypothesis, a few sidenotes should be mentioned. First, the coefficients are not BLUE since not all OLS assumptions held as described in Section 3.3, indicating that only the association and the strength of the association between variables can be interpreted. Second, all coefficients of the independent variables are based on the logarithm of the average property price; therefore, interpreting is based on the exponent of the coefficient. If not otherwise stated in the text, a coefficient of for example 0.227 indicates that a one-unit increase in the independent variable is associated with 25.5% (($\exp^{0.227}$-1) *100%) higher average WOZ values in the neighbourhood, this holds for Table 3, Table 5, Table 6, and Table 7. Third, the average property prices are used as a synonym for the average WOZ value. Fourth, the value of the explained variance in the regression models, the R^2, increases as the models progress as indicated in Table 3. The first four models (models 1, 2a, 2b, and 3) have a low explained variance and are not representative for interpreting the results. These models are mainly presented for observing the effect of adding more variables to the regression model. The final three regression models (models 4, 5, and 6) have a high explained variance of more than 80%. Therefore, these models are considered to be the relevant regression models and are used for interpreting the results in this chapter. Since model 6 has an R^2 of more than 95%, it is considered the most relevant regression model.

The variables relating to the employment sectors in the regression models indicate the numbers of employees per sector per neighbourhood.[5] For interpreting the results, it is helpful to include the share of employees per sector per neighbourhood as well. Therefore, Appendix 8 includes the descriptive statistics of the employment sectors for the number and share of employees per sector per neighbourhood. This indicates that sector OPQ 'government, education and healthcare' has the largest average share per neighbourhood of 25.5%, whereas the smallest share is from sector A 'agriculture, forestry and fishing' with 0.2%. In

[5] Note that the number of employees per sector per neighbourhood is chosen over the share of employees per sector per neighbourhood. By using the share per neighbourhood within a regression model, one employment sector must serve as reference variable. The interpretation of the other employment sectors is otherwise always compared to another employment sector. This is not representable for interpreting the association of one specific employment sector on the average property price in the Amsterdam neighbourhoods.

addition, the number of employees is divided by 100 in the dataset to enable easier interpretation of the coefficients, this applies to all regression models in Table 3. Namely, the influence of a single employee per sector on the average property price in a neighbourhood is negligible. By dividing the value of these variables by 100, the values of the coefficients and standard errors remain equal – only the location of the comma changes. The same is applicable for percentages, such as having a share of multifamily homes with a value of 0.85 (or 85%). A one-unit increase results in a value of 1.85, or in other words to an increase of 117%. Since this is not a useful number for interpreting the results, it is possible to divide the value by 100 to simulate a 1% additional share. For the coefficients and standard errors, again, only the comma changes. The results of the regression models are presented in Table 3 and for the sake of simplicity, only the relevant variables for the hypotheses are presented. The complete table of the regression models is revealed in Appendix 10.

4.1 Hypothesis 1

H_0: There is no positive association between average property prices and the presence of commercial, education, healthcare, and leisure employment sectors.

H_1: There is a positive association between average property prices and the presence of commercial, education, healthcare, and leisure employment sectors.

In earlier literature, the commercial, education, healthcare, and leisure employment sectors correlate positively with property prices. As indicated by the results in Table 3, the coefficients of just two employment sectors (sector OPQ and sector RSTU) deviate significantly from zero at the 90% confidence interval in the relevant regression models. Earlier literature suggested more convincing results by analysing the association between employment sectors and residential property prices. To substantiate this matter, Appendix 9 visualises the correlation between each employment sector and the average property price per neighbourhood using a scatterplot and a local polynomial line with a 95% confidence interval. Note that the Y-axes of these plots and graphs has a different range. Evaluating these plots and graphs does not reveal any specific trends.

According to Cao and Cory (1982), Koster and Rouwendal (2012), Lafferty and Frech (1978), and Song and Knaap (2004), the presence of the commercial sector is positively associated with house prices. The commercial sector is to be interpreted broadly and covers retail (partly sector GI), communication and information (partly sector HJ), financial services and real estate (sector KL), and business services (sector MN). It appears that none of these commercial employment sectors are associated with the average property price in the neighbourhoods.

The null hypothesis is therefore not rejected, indicating that having additional employees in the commercial sector within the neighbourhood is not evidently associated with the average property values.

Crucial to note is that sector GI (trade and catering industry) and sector HJ (transport, information and communication) might represent a dichotomy within the sector. Koster and Rouwendal (2012) reported that the presence of retail is positively associated with property prices, whereas the presence of wholesale is negatively associated with property prices. Thus, information and communication relate to the commercial sector and are expected to be positively correlated with property prices. By contrast, transport relates more to industrial employment and would thus be negatively associated with property prices according to De Vor and De Groot (2011) and Koster and Rouwendal (2012). For this reason, the values within these merged sectors might be counteractive, preventing coefficients being significantly different from zero.

The education and healthcare sector are covered in sector OPQ (government, education, and healthcare) and is responsible for 25.48% of employees on average in the neighbourhood. Earlier literature suggested a positive association between property prices and the healthcare and education sectors, whereas government was suggested to expected to be negatively associated (Koster and Rouwendal, 2012). Regression model 4 is significantly different from zero at the 90% confidence interval. This implies that a one-unit increase (100 employees) in sector OPQ is associated with 0.39% higher average property prices in Amsterdam's neighbourhoods. The null hypothesis of no positive association between the average property price and the education and healthcare employment sector was thus rejected. It appears that the potential negative association of the governmental sector is outperformed by a positive association. The assumption that sector OPQ is positively associated is accepted for model 4. The significance does not hold in the subsequent models, indicating that the additional variables do not strengthen the association between the average property price per neighbourhood and sector OPQ.

The leisure sector, covered in sector RSTU (culture, recreation, and other services), is on average responsible for 11.86% of the employees in the neighbourhoods. The final model is significantly different from zero at the 90% confidence interval. The null hypothesis of no positive association between property prices and the leisure sector is thus rejected. An additional 100 employees in sector RSTU is positively associated with 1.34% higher average property prices in the Amsterdam neighbourhoods. This result complements the study of Song and Knaap (2004), who reported higher property prices in proximity to leisure activities. The

Table 3 : Regression models

VARIABLES	(1) Diversity index (DI)	(2a) + Average DI	(2b) + Weighted average DI	(3) + Employ-ment sectors	(4) + Housing charact.	(5) + Spatial charact.	(6) + Social environment
Diversity index	0.227***	0.115***	0.164***	0.162***	0.0350*	0.0271	0.00242
	(0.0427)	(0.0361)	(0.0412)	(0.0435)	(0.0203)	(0.0189)	(0.0136)
Average diversity index		0.317***		0.298***	0.164***	0.104***	0.0740***
		(0.0448)		(0.0472)	(0.0351)	(0.0241)	(0.0201)
Weighted average diversity index			0.185***				
			(0.0451)				
Employees sector A (x100)				0.862*	-0.0610	1.095	0.596
				(0.479)	(0.294)	(0.737)	(0.560)
Employees sector BCDEF (x100)				-0.0711***	-0.0221	-0.0131	-0.00741
				(0.0235)	(0.0164)	(0.0123)	(0.00627)
Employees sector GI (x100)				-0.000783	-0.00187	-0.00256	-0.00221
				(0.00489)	(0.00301)	(0.00188)	(0.00205)
Employees sector HJ (x100)				-0.0361*	-0.00278	-0.00568	0.00386
				(0.0206)	(0.00801)	(0.00703)	(0.00507)
Employees sector KL (x100)				0.00700	0.00135	0.00118	0.000950
				(0.00536)	(0.00370)	(0.00265)	(0.00127)
Employees sector MN (x100)				-0.00610	0.00134	-0.000351	-0.00324
				(0.00980)	(0.00490)	(0.00386)	(0.00255)
Employees sector OPQ (x100)				-0.00357	0.00392*	0.00186	0.00346
				(0.00382)	(0.00214)	(0.00239)	(0.00251)
Employees sector RSTU (x100)				0.0184	0.00569	0.00588	0.0138*
				(0.0219)	(0.01000)	(0.00690)	(0.00699)
Share multifamily homes					0.172**	-0.139**	-0.0587
					(0.0707)	(0.0588)	(0.0455)
Control variables							
Structural characteristics	-	-	-	-	Yes	Yes	Yes
Spatial characteristics	-	-	-	-	-	Yes	Yes
Social environment	-	-	-	-	-	-	Yes
Observations	382	382	382	382	382	382	382
R-squared	0.248	0.395	0.317	0.440	0.824	0.898	0.951

Note: The dependent variable is the log transformed average property price per neighbourhood. The standard errors are clustered on a greater neighbourhood scale (wk_code), adding up to 96 neighbourhoods. Robust standard errors in parentheses. All models include a constant and error term. *** p<0.01, ** p<0.05, * p<0.10. Variable p_mgezw (share of multifamily homes) is a dummy, whereas the reference category is p_1gezw (share of single-family homes). The complete table including control variables is represented in Appendix 10.

presence of culture and recreation is reflected in places where people spent their leisure time, thus having these amenities close by is associated with higher average property prices.

4.2 Hypothesis 2

H_0: There is no negative association between average property prices and the presence of agriculture, industry, wholesale, and government employment sectors.

H_2: There is a negative association between average property prices and the presence of agriculture, industry, wholesale, and government employment sectors.

In earlier literature, the presence of the agriculture, industry, wholesale, and government employment sectors are expected to negatively correlate with property prices. Similar to the first hypothesis, almost none of the coefficients for the employment sectors are significantly different from zero in the relevant regression models, as presented in Table 3. Viewed together with the plots and graphs in Appendix 9, no obvious association is found between the employment sectors and average property price in the neighbourhoods.

The agricultural sector is covered by the variable of sector A as described in Appendix 3. Lafferty and Frech (1978) reported a negative association between property prices and the agricultural sector. In Amsterdam, this sector has no employees in 330 out of the 382 neighbourhoods, while the maximum number of employees per neighbourhood is 31, assumably because of the high urban level of Amsterdam. Besides not having significant values in the relevant regressions models for sector A as presented in Table 3, this sector is also less representative in Amsterdam due to the low number of observations. Regarding the industrial sector, belonging to sector BCDEF (industry and energy), it is accountable for 5.33% on average of employees within the neighbourhoods. Previous literature has reported ambiguous results. The studies of De Vor and De Groot (2011) and Koster and Rouwendal (2012) have suggested a negative association between property prices and the industry sector, whereas Cao and Cory (1981) suggested a positive association. The hypothesis is based on studies focussing on the Netherlands, and both indicate a negative association between industry and property prices. Since the results in Table 3 are not significantly different from zero, the null hypothesis is not rejected, thereby neither supporting nor disproving existing literature.

The wholesale sector is covered by sector GI (wholesale, retail, and catering industry). Since the commercial sector was expected to be positively correlated with property prices, sector GI (and thus the wholesale sector) is already covered in the discussion of the first hypothesis. The same holds for the government sector, which is covered by sector OPQ (government, education, and healthcare) in the previous hypothesis.

Several probable reasons exist for the relatively low number of significant variables in the first and second hypothesis in comparison with existing literature. First, the neighbourhood scale for measuring the association between employment sectors and property prices might have been too large, preventing variables becoming significant. Second, measuring the association of an increase in the number of employees might not be representative enough since some businesses have many employees. Third, some employment within the merged sectors might contain contradicting values. Fourth, the association between property prices and employment sectors might have a nonlinear relationship. The final potential reason is that residents in areas of different population densities might value the presence of certain employment sectors differently. Section 4.6 elaborates further on these potential reasons preventing significant variables and tries to provide solutions for these reasons.

4.3 Hypothesis 3

H_0: The association between the average property prices and the presence of employment is not stronger for multifamily homes compared to single-family homes.

H_3: The association between the average property prices and the presence of employment is stronger for multifamily homes compared to single-family homes.

Based on earlier literature, in areas where business activities are mixed with housing, multifamily homes are expected to be positively associated with property prices within the neighbourhood compared with single-family homes (Cao and Cory, 1981; Koster and Rouwendal, 2012; Song and Knaap, 2004). The reported reason is that residents of single-family homes value a homogenous residential neighbourhood more. In this study, the coefficients varied in each regression model. Regression models 4 and 5 are both significantly different from zero at the 95% confidence interval in Table 3; thus, both rejecting the null hypothesis. An ambiguous result is that the sign is positive in regression model 4, indicating that an increase in the share of multifamily homes instead of single-family homes is associated with higher average property prices in the neighbourhood. By contrast, model 5 has an opposite sign and model 6 is not significant at all. Based on the outcomes of these coefficients, a clear statement of this hypothesis is not evident.

Regarding the outcome of these coefficients, it is not clear how the average property price in the neighbourhood is related to single- or multifamily homes. Whereas model 4 supports earlier research, the R^2 value is higher in model 5. Since model 5 does not support earlier literature, this might as well be true for Amsterdam. A possible substantiation is that single-family homes generally have a greater floor space compared with multifamily homes and therefore sell for higher prices. Another substantiation for this point of view is that the Netherlands in general has an overheated housing market, which is also true for Amsterdam (Couzy and Damen, 2018). Moreover, the Randstad (Amsterdam, Rotterdam, The Hague, and Utrecht) experienced a strong growth of inhabitants in comparison with other Dutch cities from 2005 to 2020 (CBS, 2021). Due to the relatively scarce supply of available housing in the market and the large number of interested buyers, asking prices might often have been outbid. Although, this pronunciation must be handled with caution since the WOZ values lag current developments and are based on the property prices in the previous year.

In terms of the housing market, Amsterdam is a dense city and has a relatively small stock of single-family homes (only 15.5%), as presented in the descriptive statistics, Appendix 7. The neighbourhoods in Amsterdam, including all neighbourhoods with single-family housing, have a wide variety of employment sectors within their boundaries. The combination of the overheated housing market, density of the city, and low share of single-family homes therefore might offer a possible explanation for why single-family homes are not associated with lower average property prices within neighbourhoods. Residents who would like to live in homogenous neighbourhoods might not even consider living in Amsterdam and choose to live in somewhat smaller cities. Amsterdam residents might, on the other hand, prefer living in single-family homes and having the amenities of the city close by.

4.4 Hypothesis 4

H_0: There is no positive association between average property prices and a greater diversity of employment sectors.
H_4: There is a positive association between average property prices and a greater diversity of employment sectors.

The diversity index of the neighbourhood itself is significantly different from zero in the first five out of seven regression models as represented in Table 3. The descriptive statistics, Appendix 7, indicate that the minimum value of the diversity index is 1.21 and the maximum value is 6.38; Appendix 2b visualises the diversity index per neighbourhood. As described in Section 3.2, the maximum value for the diversity index by Koster and Rouwendal (2012) was reported to be 6.28, indicating that these maximum values are rather similar. The significance decreases in each successive model, whereas the explained variance becomes higher simultaneously. Model 4 is significant at the 90% confidence interval, whereas a one-unit increase in the diversity index is associated with 3.56% higher average property prices in the neighbourhood. This indicates that a more diversified neighbourhood is likely to be preferred by residents since this is associated with higher property prices. When variables relating to spatial characteristics and the social environment were added to the regression models, the coefficients of the diversity index are no longer significant. This indicates that the association between the diversity index and the average property prices in the neighbourhood is not strong enough to be significant in the presence of these additional variables.

The other variant of the diversity index translates into the average diversity index of the surrounding neighbourhoods as indicated in Table 3. This is an interesting variable since all models exhibited significant results at the 99% confidence interval. This indicates that the null hypothesis – no positive association between the average property prices and a greater diversity of employment sectors in the neighbourhood – should be rejected. This implies that neighbourhoods do not solely depend on the characteristics of their ‘own’ neighbourhood but also on the characteristics of the surrounding neighbourhoods. In this situation, the characteristics translate into employment sectors, which also translate into amenities. To account for similar characteristics in adjacent neighbourhoods, or locational similarity, the models use clustered standard errors. The model with the highest explained variance is the final model, which is used to express the association between property prices and having diverse employment in surrounding neighbourhoods. A one-unit increase of the diversity index in all surrounding neighbourhoods is associated with higher property prices in the neighbourhoods itself, on average 7.68% higher. This is a considerable increase and directly demonstrates how surrounding neighbourhoods are related to the price of their ‘own’ neighbourhood.

In the final two regression models in Table 3, the diversity index of the neighbourhood itself is not significant, whereas the average diversity index of surrounding neighbourhoods is significant in each regression model. A feasible explanation is that the association of one’s ‘own’ neighbourhood is not strong enough and is partly dependent on the situation of surrounding neighbourhoods. In addition, a one-unit increase in the average diversity index of the surrounding neighbourhoods means that all neighbourhoods increase by one-unit in the diversity index. The association is therefore stronger compared with the diversity index of the neighbourhood itself. Residents benefit from having a wide range of amenities – namely employment sectors – close by. Thus, residents do not have to travel far to obtain necessities of daily life. The rejection of the null hypothesis is closely related to the 15-Minute City. Moreno et al. (2021) mainly envisioned added value of living for residents since they spend less time travelling and have all their basic needs within an acceptable distance. The results of this study confirm those authors’ vision in terms of higher average property prices in neighbourhoods.

4.5 Other variables of interest

The hypotheses indicated the associations between average property prices in the neighbourhoods of Amsterdam with employment sectors, single- and multifamily homes, and the diversity index. Besides these variables, each regression model is extended by adding more variables, the control variables. Some of these variables were significantly different from zero, indicating that their presence is associated with housing prices. The table including the results of all variables in the regression models is presented in Appendix 10. This section highlights a few interesting variables, these are the variables: the number of homes, the floor space and the construction year. In Appendix 11 is elaborated further on the remaining significant variables, these are the variables with the share of: owner-occupied and rental properties; vacant and inhabited homes; water and land bodies; average address density; Western and non-Western immigrants; low- and high-income households; households below social minimum; and households that receive government benefits. Besides these variables containing shares, there is also elaborated further on the variables with distance to: education; transfer- and train stations; swimming pools; ice rinks; solariums; and museums.

Housing characteristics

An interesting variable is the number of homes (*homes*) within a neighbourhood. Regression model 6 in Appendix 10, the most relevant model, exhibits a significant coefficient at the 99% confidence interval. Increases of 1, 10, and 100 home(s) within a neighbourhood are associated with lower property prices, on average 0.013%, 0.131%, and 1.301% lower, respectively. Thus, it can be stated that Amsterdam's residents do not prefer the presence of additional housing within their own neighbourhood.[6]

The floor space variables are interpreted as dummy variables, whereas the reference category is the smallest floor space, ranging from 0 to 50 square metres. Within the dataset, the share of homes within each category is given in percentages. All variables exhibit significant results at the 99% confidence interval. A clear and logical pattern is visible in the share of the floor spaces: the greater the floor space, the higher the property prices. To interpret the association of 1% higher property values of these variables, the coefficients are divided by 100. Table 4 represents the magnitude of the other dummy variables. Surprisingly, the difference between

[6] A notable aspect is that model 4 and model 5 do not show significant values for the number of homes. The rule of thumb for multicollinear values is that the Variance Inflation Factor (VIF) must not exceed the value 10 or that the correlation between independent variables do not exceed 0.8. No variables in the regression models do exceed this number. Despite, the VIF for the variable of the number of homes (*homes)* is considerably different between the regression models. The VIF in model 4 is 1.92, in model 5 is 2.46 and in model 6 is 9.51. This indicates that the variable *homes* in model 6 almost reaches the upper limit for being multicollinear. For this reason, there might be a chance that this variable is still correlated to other variables.

Table 4 : Magnitude of floor space in relation to property prices

Variable	***Coefficient /100***	***Magnitude***
Floor space 50 m^2 or smaller (reference category)	-	-
Floor space 51-75 m^2	0.00294	0.29%
Floor space 76-100 m^2	0.00490	0.49%
Floor space 101-150 m^2	0.00521	0.52%
Floor space 151-250 m^2	0.01200	1.21%
Floor space 251 m^2 or larger	0.02586	2.62%

Note: The coefficients correspond with model 6 in Appendix 10. The dependent variable is the average property price, and the given variables are dummy variables, whereas the reference category is the floor space 50 square metres or smaller. If the share of floor space from 51 to 75 metres is 1% higher, the average property price in the neighbourhood is associated with 0.29% higher property prices compared with a situation where the share of floor space ranged from 0 to 50 square metres.

the share of floor surface in the ranges of 76–100 and 101–150 square metres is not that large at only 0.33%. A more gradual increase towards the next category would have been more logical since a higher floor space is often related to a higher property price, but according to the data, this does not hold. A 1% higher share of properties with a floor surface larger than 150 square metres even is associated with more than a 1% higher average property prices in the neighbourhood, compared with a situation where the share of the floor space ranges from 0 to 50 square metres.

In addition, a pattern is found for the dummy variable of construction year and the reference category is properties constructed after 2000. In each of the regression models, at least one of the construction year variables is significantly different from zero. Notably, however, the average value of properties within the neighbourhood is associated with the lowest average property prices when the share of properties constructed after World War II increases, being significantly different from zero at the 90% confidence interval in regression models 5 and 6 in Appendix 10. After World War II, the Netherlands experienced a great shortage of homes, and therefore, more than one million were built in less than 15 years according to Broekhoven (2017). As this author reported, these buildings today are often outdated and due for renewal. These underlying reasons seem to be confirmed when examining the results of the present study. The novelty of properties constructed after the year 2000 slowly fades, what reflects in lower housing prices for properties constructed between the period 1980 and 1999. Although the coefficients of the property construction in the year 1980 and 1999 have negative signs, none of the coefficients are significantly different from zero. Therefore, no clear statement of having a lower association with these property types hold. On the other hand, properties built before World War II have higher values than post-2000 properties and are significantly different from zero at the 99% confidence interval in regression model 4 and 5. The obvious reason is that those properties bear historical value and are located in proximity to the centre, which is reflected in an association with higher property prices.

4.6 Robustness test 1: Businesses instead of employees

To check whether the assumptions made in the analysis are true, robustness tests are conducted. By creating some variations in the dataset or changing the setup, this study checks whether the results remain similar to the earlier results. Specifically, the robustness is checked by performing four tests as described in the sensitivity analysis in Section 3.5.

The first robustness test uses businesses instead of employees. The employment sectors in this research are interpreted according to the number of employees per sector per neighbourhood. The descriptive statistics of these variables are presented in Appendix 8. Some businesses have more than a thousand employees, whereas other businesses are self-employed. To factor out the impact of these large companies within the dataset, the number of businesses is taken into account. Regression models 1, 2a, and 2b are irrelevant since these models do not use the variables of the employment sectors at all. In addition to that, only the coefficients that are significantly different from zero are elaborated on.

Similar to the regression models in Table 3, the explained variance in model 3 is too low to interpret the results. The average diversity index of surrounding neighbourhoods remains significant in all the relevant models with an R^2 higher than 80%. When evaluating the R^2, the values in Table 5 are quite similar to the values in Table 3. A one-unit increase in the diversity index when using businesses instead of employees remains quite similar. The maximum deviation occurs in regression model 4, whereas the coefficient of the average diversity index is 0.017 lower compared with the results in Table 3. Therefore, it can be stated that the association of measuring the number of businesses or the number of employees in the employment sectors does not vary much when considering the average diversity index of the surrounding neighbourhoods.

Regarding regression models 4, 5, and 6 in Table 5, the employment sectors with coefficients being significant from zero are sectors GI (trade and catering industry), HJ (transport, information, and communication), MN (business services), OPQ (government, education, and healthcare), and RSTU (culture, recreation, and other services). Compared with Table 3, more coefficients within the employment sectors are significant, indicating that the number of businesses could explain the variation in average property prices within the neighbourhoods better.

Within a neighbourhood, an increase in the number of businesses in sector GI is significantly associated with lower property prices at the 95% and 90% confidence interval in regression

Table 5 : Regression models with businesses instead of employees

VARIABLES	(3) + Employ- ment sectors	(4) + Housing charact.	(5) + Spatial charact.	(6) + Social environment
Diversity index	0.0702* (0.0370)	0.00881 (0.0196)	0.00981 (0.0132)	-0.00217 (0.0125)
Average diversity index	0.206*** (0.0440)	0.147*** (0.0339)	0.0949*** (0.0220)	0.0770*** (0.0186)
Businesses sector A	0.0306*** (0.00733)	0.00131 (0.00522)	0.0231 (0.0150)	0.0125 (0.0106)
Businesses sector BCDEF	-0.00140 (0.00190)	-0.000972 (0.00113)	-0.000630 (0.000841)	-0.000905 (0.000638)
Businesses sector GI	-0.000666 (0.000582)	-0.000370 (0.000234)	-0.000442** (0.000193)	-0.000374* (0.000189)
Businesses sector HJ	-0.00668*** (0.00165)	0.000237 (0.00121)	-0.000981 (0.000745)	0.00115** (0.000571)
Businesses sector KL	0.00233 (0.00286)	-7.87^e-05 (0.00137)	8.13^e-05 (0.000999)	-0.000142 (0.000623)
Businesses sector MN	0.00345*** (0.000587)	0.000929** (0.000455)	0.000851** (0.000344)	-0.000201 (0.000296)
Businesses sector OPQ	0.000783 (0.00124)	0.000688 (0.000752)	0.000880* (0.000503)	0.000419 (0.000370)
Businesses sector RSTU	-0.000501 (0.000837)	0.000459 (0.000488)	0.000241 (0.000394)	0.000879*** (0.000292)
Control variables				
Structural characteristics	-	Yes	Yes	Yes
Spatial characteristics	-	-	Yes	Yes
Social environment	-	-	-	Yes
Observations	382	382	382	382
R-squared	0.546	0.834	0.904	0.952

Note: The dependent variable is the log transformed average property price per neighbourhood. The standard errors are clustered on a greater neighbourhood scale (wk_code), adding up to 96 neighbourhoods. Robust standard errors in parentheses. All models include a constant and error term. *** p<0.01, ** p<0.05, * p<0.10.

models 5 and 6, respectively. Since the explained variance is higher in regression model 6, this model is used for evaluation. An increase of one additional business in the trade and catering industry per neighbourhood is associated with 0.04% lower average property prices in Amsterdam neighbourhoods. As mentioned in Section 4.1, the employment within this merged sector might be contradictory; retail is expected to be positively associated with property prices, whereas wholesale was expected to be negatively associated with property prices (Koster and Rouwendal, 2012). Based on the results in Table 5, the trade and catering industry is negatively associated with property prices overall.

In regression model 6, an increase in the number of businesses in sector HJ (transport, information, and communication) is positively associated with average property prices within the neighbourhoods at the 95% confidence interval. An increase of one business is associated with 1.16% higher average property prices. Koster and Rouwendal (2012) reported a positive association between property prices and the commercial sector. Sector HJ tends mostly towards the commercial sector, and therefore, the results in Table 5 are the opposite to those reported in earlier literature. However, the employment in sector HJ also contains transport and storage, which are associated more with industrial sites. According to earlier literature, this should be associated more with lower property prices (De Vor and de Groot, 2011; Koster and Rouwendal, 2012). Therefore, the values within this merged sector might work counteractively, but overall, this merged sector is negatively associated with property prices.

Considering the relevant regression models, sector MN (business services) is significantly different from zero at the 95% confidence interval in regression models 4 and 5. Business services cover, for example, consultancy, research, movable property rental, and other business services. An increase of one business in sector MN within the neighbourhood is associated with higher average property prices in the neighbourhood, on average being 0.09% higher. This positive association supports earlier literature since the presence of commercial activities is related to higher property prices (Cao and Cory, 1981; Koster and Rouwendal, 2012; Song and Knaap, 2004). Although the number of employees for this sector is not associated with average property prices in the neighbourhood according to Table 3, the number of businesses is associated with the average property prices.

Sector OPQ (government, education, and healthcare) is significant at the 90% confidence interval in regression model 5. Compared with the previous variables, the indication of significance is weaker since only one coefficient out of the three relevant regression models is significant, and moreover, the significance is lower. A one-unit increase in the number of businesses in the government, education, and healthcare sector is associated with 0.09% higher average property prices in the neighbourhoods. This supports the study of Cao and Cory (1981), who reported a positive association with public land use. In addition, this partly supports the study of Koster and Rouwendal (2012), who reported a positive association with the education and healthcare sector but a negative association with the governmental sector. Again, based on earlier literature, these sectors within OPQ might have counteractive values, but the results indicate that the merged sector does indeed is positively associated with the average property prices in Amsterdam.

The final employment sector RSTU (culture, recreation, and other services) are significantly different from zero at the 99% confidence interval in regression model 6. One additional business in sector RSTU is associated with higher average property prices in the neighbourhood, on average 0.09% higher. Table 3 also revealed a significant coefficient for this employment sector in regression model 6. This strengthened the hypothesis that recreation and culture are positively associated with average property prices.

4.7 Robustness test 2: 21 employment sectors instead of 8

In previous models, the employment sectors were merged from 21 to 8 sectors according to the CBS categories. Appendix 3 describes all of these individual employment sectors and the descriptive statistics are presented in Appendix 8. According to Table 3, almost none of the variables resulted in significantly different coefficients, indicating that these employment sectors are not strongly associated with the average property prices in Amsterdam's neighbourhoods. It is therefore useful to consult the 21 sectors as single entities and check whether these single entities significantly correlate with property prices. Table 6 indicates the coefficients of the individual employment sectors. The table uses the same structure as Table 5. Similarly, regression model 3 is not considered relevant, the R^2 values are quite similar compared to Table 3, and there is only elaborated on the significant coefficients.

The first variables associated with the average property price in the neighbourhood are sector B (mining and quarrying), sector D (electricity, gas, steam, and air conditioning supply), and sector E (water supply, sewerage, waste management, and remediation activities). Appendix 8 indicates that the average share of each of these individual employment sectors in the neighbourhood is less than 0.2%. Considering the data, sectors B, D, and E were only represented by employees living in 53, 35, and 45 out of 382 neighbourhoods, respectively. These numbers are considerably lower compared to other employment sectors. Although the value of many of the observations is zero, interpreting these coefficients remains valid since all observations are included.

The coefficients of sector B are considerably high compared to the other coefficients in Table 6 and significantly different from zero in model 4 and 5. An additional 100 employees in sector B is associated with 6.3 times higher property prices when considering model 5. The mean share of sector B is on average lower than 0.1% according to Appendix 8. Considering earlier literature, the industry sector was expected to be negatively associated with property prices and therefore the outcomes of the coefficients are contradictory with earlier literature (De Vor and De Groot, 2011; Koster and Rouwendal, 2012). A possible explanation for the high value

Table 6 : Regression models with 21 instead of 8 employment sectors

VARIABLES	(3) + Employ- ment sectors	(4) + Housing charact.	(5) + Spatial charact.	(6) + Social environment
Diversity index	0.165*** (0.0333)	0.0489*** (0.0186)	0.0364** (0.0165)	0.0132 (0.0125)
Average diversity index	0.244*** (0.0375)	0.144*** (0.0261)	0.0900*** (0.0229)	0.0655*** (0.0192)
Employees sector A (x100)	1.053** (0.411)	0.0699 (0.259)	1.212 (0.737)	0.624 (0.521)
Employees sector B (x100)	-0.245 (1.786)	3.221** (1.516)	1.990* (1.153)	0.406 (0.577)
Employees sector C (x100)	-0.154*** (0.0432)	-0.0205 (0.0212)	0.0393 (0.0280)	-0.0190 (0.0222)
Employees sector D (x100)	0.0474 (0.0421)	-0.0453* (0.0262)	-0.0259 (0.0203)	-0.0196 (0.0150)
Employees sector E (x100)	-0.0572*** (0.0143)	0.0202* (0.0122)	0.0101 (0.0124)	0.00752 (0.00912)
Employees sector F (x100)	-0.160** (0.0805)	-0.0693*** (0.0256)	-0.0574*** (0.0192)	-0.0199 (0.0147)
Employees sector G (x100)	-0.00847 (0.00934)	-0.00642 (0.00479)	-0.00650** (0.00292)	-0.00510** (0.00259)
Employees sector H (x100)	-0.00534 (0.0220)	0.00372 (0.0121)	0.00133 (0.0111)	0.0117** (0.00458)
Employees sector I(x100)	0.00777 (0.0106)	0.00771 (0.00762)	0.00175 (0.00628)	0.00563 (0.00464)
Employees sector J (x100)	-0.0201 (0.0177)	-0.00551 (0.00946)	-0.0109 (0.00852)	-1.50e-05 (0.00499)
Employees sector K (x100)	-0.0125* (0.00742)	0.00305 (0.00468)	0.00195 (0.00371)	0.00257 (0.00244)
Employees sector L (x100)	0.230*** (0.0701)	0.0588 (0.0368)	0.0442 (0.0281)	0.00993 (0.0180)
Employees sector M (x100)	-0.0176 (0.0145)	-0.00608 (0.00896)	-0.00490 (0.00833)	-0.00448 (0.00402)
Employees sector N (x100)	-0.00927 (0.0167)	0.00284 (0.0129)	0.00128 (0.00697)	-0.00103 (0.00447)
Employees sector O (x100)	-0.0399*** (0.0154)	0.00443 (0.00598)	-0.00177 (0.00501)	-0.00166 (0.00506)
Employees sector P (x100)	0.0111 (0.0123)	0.00652 (0.00540)	-6.32e-06 (0.00498)	-0.00462 (0.00422)
Employees sector Q (x100)	-0.00224 (0.00456)	0.00367 (0.00335)	0.00336 (0.00306)	0.00691** (0.00271)
Employees sector R (x100)	0.0622** (0.0242)	0.0145 (0.0113)	0.0111 (0.00745)	0.0189** (0.00808)
Employees sector S (x100)	-0.0850* (0.0494)	-0.0100 (0.0298)	0.0104 (0.0220)	0.00487 (0.0156)
Employees sector U (x100)	1.666** (0.647)	0.0185 (0.306)	0.125 (0.267)	0.117 (0.239)
Control variables				
Structural characteristics	-	Yes	Yes	Yes
Spatial characteristics	-	-	Yes	Yes
Social environment	-	-	-	Yes

TABLE CONTINUED

Observations	382	382	382	382
R-squared	0.507	0.833	0.903	0.954

Note: The dependent variable is the log transformed average property price per neighbourhood. The standard errors are clustered on a greater neighbourhood scale (wk_code), adding up to 96 neighbourhoods. Robust standard errors in parentheses. All models include a constant and error term. *** p<0.01, ** p<0.05, * p<0.10.

for the coefficients is that the mining and quarrying businesses are more likely to locate in the periphery of Amsterdam and therefore located in neighbourhoods with large homes with higher property prices, resulting in a large value for the coefficient.

The coefficients of sector D and sector E are only significant at the 90% confidence interval in regression model 4 in Table 6. The coefficient of sector D indicates a negative association with the property prices in the neighbourhood and sector E indicates a positive association with property prices. Since these sectors are mostly related to the industry sector as well, sector D supports earlier literature and sector E does not support earlier literature (De Vor and De Groot, 2011; Koster and Rouwendal, 2012). An important notice is that the association is not that strong since only one out of the three relevant regression models is significant, and the significance level is at the 90% confidence interval. The subsequent significant variables (sectors F, G, H, R, and Q) do have employees in almost all neighbourhoods – only 12 neighbourhoods do not have employees in these individual sectors. Sector F (construction) is negatively associated with the average property price in the neighbourhood considering the relevant regression models 4 and 5. Within the construction sector, each company relating to the construction of buildings, roads, civil engineering, underground installations, and hydraulic engineering is covered. An increase of 100 employees working in construction within a neighbourhood is associated with lower property prices, being on average 5.58% lower. Earlier literature has not focused on the construction industry specifically, but it can be assumed that this belongs to the industrial land use. This supports earlier literature that has found industrial land use in proximity to homes to be negatively associated with property prices (De Vor and De Groot, 2011; Koster and Rouwendal, 2012).

Sector G (wholesale and retail trade and repair of motor vehicles) also results in significant coefficients at the 95% confidence interval in regression models 5 and 6. The share of this sector per neighbourhood accounts on average for 12.7% as presented in Appendix 8. The coefficients indicate a negative association between sector G and the average property prices in Amsterdam's neighbourhoods. Still, this employment sector as a single entity might have contradicting values. According to Koster and Rouwendal (2012), property prices are valued negatively in proximity to wholesale, whereas properties in proximity to retail are valued positively. The negative association outperforms the potential positive association. Model 5 and 6 both indicate a significantly negative association with property prices by an increasing number of employees in sector G, being 0.51% lower on average.

Another sector associated with average property prices in Amsterdam's neighbourhoods is sector H (transportation and storage). The coefficient is significantly different from zero at the 95% confidence interval in regression model 6. An additional 100 employees in this sector is associated with 1.18% higher average property prices in the neighbourhood. This supports the findings in Table 3, where an increase in the number of businesses in the merged sector HJ (transport, information, and communication) was also associated with higher property prices. Nevertheless, these findings do not support earlier literature since most storage and transportation businesses are located in more industrial areas. This is expected to be negatively associated with property prices according to De Vor and De Groot (2011) and Koster and Rouwendal (2012); however, this does not hold for Amsterdam.

In Table 3, the merged sector OPQ (government, education, and healthcare) was significantly different from zero and positively associated with average property prices in the neighbourhoods. Table 6 indicates that not all of these individual employment sectors are significantly different from zero; only sector Q (human health and social work activities) in regression model 6. This sector covers, for example, hospitals, physiotherapists, dentists, nursing facilities, youth services, and social services. This sector is associated with the average property price at the 95% confidence interval, where an additional 100 employees in the sector is associated with 0.69% higher average property prices in the neighbourhood. This implies that sectors O and P are not significantly associated with average property prices as being single entities. Based on these results, residents would like to have all types of healthcare facilities and social services in direct proximity. A potential explanation is that residents can more easily reach these amenities in (urgent) health-related situations.

Similar for the merged sector RSTU (culture, recreation and other services), the results as presented in Table 3 resulted in a significant positive association with average property prices in the neighbourhoods. The only employment sector as an individual entity that was significantly different from zero at the 95% confidence interval was sector R (culture, sports, and recreation) in regression model 6. The other employment sectors, S (other service activities) and U (extraterritorial organisations and bodies), are not associated with the average property prices. It can therefore be assumed that the significance in Table 3 originates mainly from sector R. An additional 100 employees in sector R are associated with higher average property prices in the neighbourhood, on average 1.91% higher. This sector is related to the public interest since people are likely to use these amenities for leisure. These results support the study of Daams et al. (2016), who reported higher property prices in proximity to recreation. The results also support the study of Jang and Kang (2015), Koster and Rouwendal (2012), and Song and Knaap (2004), who have indicated higher property prices in proximity to open and public parks, which are mostly associated with the recreational sector.

In the cultural and recreational sector, two employment sectors are significantly different from zero at the 95% confidence interval: sector Q (human health and social work activities) and sector R (culture, sports, and recreation). Both are significant in the final regression model. An additional 100 employees in sectors Q and R are associated with higher average property prices in the neighbourhood, being on average 0.69 and 1.91% higher, respectively. This sector relates to the public interest as well and supports the study of Cao and Cory (1981) who reported a positive association in property prices in the presence of public serves.

4.8 Robustness test 3: Testing nonlinearity in the employment sectors

In this study, it was assumed that employment sectors are either positively or negatively associated with property prices and have a linear relationship. It might also be true that property prices within the neighbourhoods have a nonlinear relationship with the number of employees in each employment sector. All employment sectors are squared and added as extra variables into the final four regression models. Since each of the employment sector variables is used twice, once in linear form and once in quadratic form, the coefficients of the linear employment sectors become irrelevant (Brooks and Tsolacos, 2010). Even if the coefficients of the linear employment sectors become significant in addition of the nonlinear employment sectors, they are not considered. Table 7 represents the nonlinear regression models and uses the same structure as Table 5. Similarly, regression model 3 is not considered relevant and there is only elaborated on the significant coefficients.

Regression model 3 does not exhibits an increase in explained variance and remains low with a value of 47.4% as presented in Table 7. The explained variance of regression models 4, 5, and 6 almost equalled the explained variance in Table 7, indicating that the squared terms do not have a great impact on the explained variance. Even though some of the coefficients are significantly different from zero in these regression models, these coefficients are left out of consideration. Within regression model 4, none of the quadratic employment sectors exhibit significant values. In regression model 5, only sector OPQ (government, education, and healthcare) is significantly different from zero at the 95% confidence interval. In the final regression model, sector OPQ becomes more significant from zero, and it is significant at the 99% confidence interval. For this reason, the coefficient within the final model is used for further

Table 7 : Regression models with squared employment variables

VARIABLES	(3) + Employment sectors	(4) + Housing charact.	(5) + Spatial charact.	(6) + Social environment
Diversity index	0.162***	0.0637**	0.0392*	0.0164
	(0.0341)	(0.0258)	(0.0205)	(0.0144)
Average diversity index	0.298***	0.156***	0.105***	0.0709***
	(0.0393)	(0.0260)	(0.0231)	(0.0184)
Employees sector A (x100)	0.862**	-0.254	1.627	1.338
	(0.412)	(1.212)	(1.150)	(0.851)
Employees sector BCDEF (x100)	-0.0711***	-0.0602**	-0.0207	-0.0183
	(0.0272)	(0.0285)	(0.0273)	(0.0160)
Employees sector GI (x100)	-0.000783	-0.00721	-0.00155	-0.00286
	(0.00517)	(0.00713)	(0.00527)	(0.00467)
Employees sector HJ (x100)	-0.0361**	-0.0224	-0.0134	0.00603
	(0.0174)	(0.0250)	(0.0203)	(0.0129)
Employees sector KL (x100)	0.00700	0.00465	0.00490	0.00278
	(0.00531)	(0.00797)	(0.00621)	(0.00452)
Employees sector MN (x100)	-0.00610	-0.00664	-0.00437	-0.00821*
	(0.0120)	(0.0104)	(0.00805)	(0.00483)
Employees sector OPQ (x100)	-0.00357	-0.00208	-0.00542	-0.00828**
	(0.00405)	(0.00486)	(0.00440)	(0.00363)
Employees sector RSTU (x100)	0.0184	0.0461	0.0282	0.0457**
	(0.0212)	(0.0330)	(0.0259)	(0.0196)
Employees sector A (squared)	-0.000792	3.57e-05	-0.000246	-0.000300
	(0.000570)	(0.000407)	(0.000422)	(0.000286)
Employees sector BCDEF (squared)	7.27e-07	3.52e-07	8.65e-08	1.05e-07
	(4.59e-07)	(2.30e-07)	(2.13e-07)	(1.24e-07)
Employees sector GI (squared)	5.08e-08	2.69e-08	-5.62e-09	1.11e-09
	(4.46e-08)	(2.69e-08)	(1.91e-08)	(1.54e-08)
Employees sector HJ (squared)	9.15e-07***	1.90e-07	7.04e-08	-3.33e-08
	(3.13e-07)	(2.45e-07)	(2.05e-07)	(1.10e-07)
Employees sector KL (squared)	3.77e-09	-7.56e-09	-5.32e-09	-3.11e-09
	(2.34e-08)	(9.38e-09)	(7.21e-09)	(5.24e-09)
Employees sector MN (squared)	-7.70e-09	2.81e-08	1.07e-08	1.38e-08
	(4.75e-08)	(2.61e-08)	(1.95e-08)	(1.15e-08)
Employees sector OPQ (squared)	4.18e-08**	1.67e-08	2.05e-08**	3.18e-08***
	(1.79e-08)	(1.10e-08)	(1.01e-08)	(8.09e-09)
Employees sector RSTU (squared)	-1.46e-06***	-5.32e-07	-2.92e-07	-3.97e-07**
	(5.27e-07)	(3.56e-07)	(2.63e-07)	(1.96e-07)
Control variables				
Structural characteristics	-	Yes	Yes	Yes
Spatial characteristics	-	-	Yes	Yes
Social environment	-	-	-	Yes
Observations	382	382	382	382
R-squared	0.474	0.828	0.900	0.954

Note: The dependent variable is the log transformed average property price per neighbourhood. The standard errors are clustered on a greater neighbourhood scale (wk_code), adding up to 96 neighbourhoods. Robust standard errors in parentheses. All models include a constant and error term. *** p<0.01, ** p<0.05, * p<0.10.

elaboration. In addition, sector RSTU (culture, recreation, and other services) is significantly different from zero at the 95% confidence interval within the final regression model.

These significant values indicate that the squared employment sectors have a nonlinear association with the average property price per neighbourhood. A positive coefficient indicates that the nonlinear association is U-shaped, whereas a negative coefficient indicates that the nonlinear association is an upside-down U-shape. Sector OPQ is positive and becomes a U-shape, the figure is presented in Appendix 12. First, when the number of employees increases within a neighbourhood, the prices become more negative up to 1,500 employees. When the number of employees increases even further, the association with property prices becomes positive. However, only eight neighbourhoods have more than 1,500 employees, making the nonlinear association weaker than it may appear in first instance. In addition, this indicates that residents do prefer a smaller share of the government, education, and healthcare sector within their neighbourhood, but do prefer a greater share of this sector within their neighbourhood.

For the sector of culture, recreation, and other services, an opposite trend occurs as presented in Appendix 12. When the number of employees within this sector increases, the average property price in the neighbourhood is associated with higher property prices in the neighbourhood. However, there is a turning point when the number of employees within this sector reaches approximately 500 employees. Subsequently, the average property price in the neighbourhood is associated with lower property prices, indicating that residents dislike having a large share of cultural and recreational activities within their neighbourhood. For this sector too, only six neighbourhoods have values greater than 500, again making the nonlinear association weaker than it may appear in first instance. Moreover, it is assumable that residents like the presence of some recreational and cultural activities, but that a turning point arises when these neighbourhoods become too touristic and crowded.

Considering the earlier results, sector OPQ was positively associated with property prices at the 90% confidence interval in Table 3 and Table 5. In Table 6, that uses employment sectors as single entities, specifically sector Q (human health and social activities) was positively associated and significant at the 95% confidence interval. It might therefore be true that the positive association in Table 3 and Table 5 mainly originates from the neighbourhoods with the most employees in sector OPQ, whereas sector Q might be responsible for the greatest part of the positive association. Sector RSTU was positively associated at the 90% and 99% confidence intervals in Table 3 and Table 5, respectively. Table 6 indicated a positive association with only sector R (culture, sports, and recreation) at the 95% confidence interval. For sector RSTU too, the positive association might especially originate from the neighbourhood with the greatest number of employees in sector RSTU, whereas sector R might have been responsible for the greatest part of the positive association.

4.9 Robustness test 4: Chow test on population density

A Chow test is performed to check whether differences exist among different subgroups relative to the pooled model. The pooled model is the final regression model, which is the most relevant model due to its R^2 being higher than those of the other regression models. Since many of the used variables consist of merged data, it is not possible to measure, for example, the difference between owner-occupied and rental homes since each neighbourhood contains a share of such homes. The Chow test is performed by creating subgroups for population densities. Residents of Amsterdam who live in a higher density neighbourhoods might value (some) employment sectors differently compared with those in lower density neighbourhoods. As indicated in the descriptive statistics in Appendix 7, the population density ranges between 34 and 35,903 residents/sq. km. An important notice is that population density is divided into equal groups, based on the number of observations and not based on the values of the observations. In this way, the Chow test compares groups with equal number of observations. The total of 382 neighbourhoods are divided into three groups of 127 or 128 neighbourhoods, resulting in:

- Group 1: lowest population density group, between 34 and 8924 residents/sq. km
- Group 2: middle population density group, between 8925 and 16,975 residents/sq. km
- Group 3: highest population density group between 16,976 and 35,903 residents/sq. km

To give an indication, the city with the highest population density in the Netherlands in 2018 is The Hague with an average of 6,459 residents per square kilometre, whereas Amsterdam is ranked fourth with an average of 5,160 residents per square kilometre (CBS, 2018). The reason that the groups have relatively high population densities, is that the number of residents is based on the neighbourhood scale where the average size per neighbourhoods is less than 0.5 square kilometre. Also, neighbourhoods without residents are not taken into account, resulting in a higher number of residents in the inhabited neighbourhoods. Thus, when having many apartments for example, it drives up the average population density considerably.

The null hypothesis for the Chow test is as follows: The intercepts and slopes of neighbourhoods with the lowest, middle, and highest density population groups are identical. Appendix 13 presents the calculation of the F value for the Chow test. The F value is slightly greater than the critical value at the 5% significance level. The null hypothesis is therefore rejected, indicating that the intercepts and slopes of neighbourhoods with the lowest, middle, and highest density population groups are not identical. Therefore, it can be stated that residents living in a neighbourhood with another population density value characteristics within their neighbourhood differently since the average property prices in those neighbourhoods are not identical. Appendix 14 indicates the difference between the pooled regression model and the regression models of the subgroup.

The first point to note in Appendix 14 is that the R^2 of the subgroups were higher compared with the pooled model, even though the sample sizes are smaller. This implies that all groups explain more of the variation in the average property price per neighbourhood. The most likely explanation is that the groups with similar population densities have more similar characteristics, which eventually results in higher explained variance.

Appendix 14 indicates that with a different population density, only one variable is significantly differed from zero at the 90, 95, or 99% confidence interval when considering employment sectors, namely sector A (agriculture, forestry, and fishing). It was significantly different at the 95% confidence interval with a considerably high value of 5.193. It might be the case that agricultural businesses locate in the periphery of Amsterdam and have considerably larger homes that are more expensive. Noteworthy, sector A might not be representative in the subgroups due to many observations with a value of zero; groups 1, 2, and 3 were only represented by employees in 18, 14, and 21 neighbourhoods, respectively. Brook and Tsolacos (2010) argue that at least 30 observations are required to be able to interpret data. Since each group contains all observations, no observations are missing. Despite, important to consider is that the data is based on the few groups that contain values for sector A that might indicate the relatively high value of the coefficient.

The diversity index of the neighbourhoods is also not significant in the subgroups, suggesting that population density does not explain differences in property prices if the diversity index increases. By contrast, the average diversity index of surrounding neighbourhoods did have significant results. The pooled model is significantly different at the 99% confidence interval, whereas groups 1 and 2 are significant at the 90% confidence interval, and group 3 is not significant at all. The coefficient of the average diversity of surrounding neighbourhoods in group 1 is the highest, indicating that higher average property prices in neighbourhoods with the lowest population density group is likely to be most evident. A one-unit increase in the average diversity index of surrounding neighbourhoods is associated with higher property prices within the neighbourhood, being 8.11% higher on average. For a one-unit increase in the neighbourhoods with a middle population density group, property prices are associated with 5.84% higher property prices. Neighbourhoods with high population density groups do not exhibit significant results, indicating that they are not associated with higher or lower property prices when the surrounding neighbourhoods experience an increase in the diversity index.

An increase in the number of homes (*homes*) relates negatively to property prices within neighbourhoods with the lowest and middle population density groups. The coefficient of group 2 is similar to the pooled model but significant at the 95% instead of the 99% confidence interval. The coefficient of group 1 is slightly more negative than the pooled model and significant at the 90% confidence interval. It is interesting that the neighbourhoods in the highest population density group are not significant. They already have many residents as well as homes; therefore, it can be assumed that the extra homes within a neighbourhood are not associated with the average property prices within a neighbourhood. On the other hand, there is a negative association with additional homes and property prices for the neighbourhoods in the lowest and medium population density groups: thus, these groups are associated with higher property prices when the number of homes within the neighbourhood is lower.

The final variable of interest in the Chow test is the share of owner-occupied homes (*p_koopw*). In the pooled model, the coefficient is significantly different from zero at the 95% confidence interval. Groups 2 and 3 are significantly different from zero at the 90% and 99% confidence intervals, whereas the negative association is stronger for the neighbourhoods with the highest population density groups. The average share of owner-occupied homes is a minority with 32.1% per neighbourhood compared with rental homes according to Appendix 8. The reason rental homes are valued higher than owner-occupied homes is not clear. Thus, having a higher share of owner-occupied homes, and thus a lower share of rental homes, in the neighbourhood is associated with lower property prices in group 2 and 3, and vice versa.

5. CONCLUSION AND FUTURE RESEARCH

This chapter consists of three sections. First, the conclusion and discussion of this study are presented, followed by the policy relevance and finally by recommending future research.

5.1 Conclusion and discussion

The purpose of this research was to empirically investigate which employment sectors are negatively or positively associated with property prices and whether a wide variety of employment sectors is related to higher property prices. According to existing literature, property prices are determined by many factors. The presence of certain employment sectors and a variety of employment sectors are not only responsible for the variation in property prices. Also, housing characteristics (e.g., floor area and construction year), spatial characteristics (e.g., distance to public transport or certain amenities) and the social environment (e.g., population density or the share of low-income households) determine variations in the property prices. This research attempted to fill the gap in existing literature by measuring the association between residential property prices and having a wide variety of employment sectors within and beyond the neighbourhood. This variety was measured using a diversity index. The method adopted involved a multiple linear regression model; the average property price per neighbourhood was the dependent variable, the independent variables were the diversity index and eight different employment sectors, and the control variables comprised the housing, spatial, and social environmental characteristics of the neighbourhoods. Amsterdam served as the case study; specifically, Amsterdam's 382 neighbourhoods were used with residential properties and businesses.

The first empirical results suggest that out of the three relevant regression models, all of which consisted of eight employment sectors, only two coefficients were significantly different from zero at the 90% confidence interval. The first one was from sector OPQ (government, education, and healthcare), which was positively associated with average property prices in the neighbourhood. This partly supports earlier literature since healthcare and education were expected to be positively correlated but government was expected to be negatively correlated with property prices. The second significant coefficient was from sector RSTU (culture, recreation, and other services) and was positively associated with the average property price. Earlier literature suggested a positive association with recreation and leisure activities; therefore, the empirical results support earlier literature. Despite the few significant coefficients and relatively low significance, the empirical results are not entirely convincing and variations in the employment sectors were found in robustness tests to either support or contradict the results.

The other independent variable of interest was the diversity index. A diversity index with a value close to 1 indicated no diversity in the neighbourhood, whereas a higher value indicated high diversity in the neighbourhood. The values for this study ranged between 1.2 and 6.4. An increase in the diversity index of a neighbourhood was not strongly positively associated with property prices in the neighbourhood. Out of the three relevant regression models, only the model with the lowest explained variance was significantly different at the 90% with a positive sign for the coefficient. Existing literature suggested that more diverse employment within close proximity is related to higher property prices, this study is not strongly supporting these results on the neighbourhood scale. On the other hand, this study measured whether surrounding neighbourhoods are associated with the average property price within a neighbourhood. For each of the neighbourhoods, a value for the average diversity index of surrounding neighbourhoods was calculated. When this value was higher, there was a strong association with higher average property prices in the neighbourhood since all relevant regression models were significantly different from zero at the 99% confidence interval. Considering the most relevant model with the highest explained variance, a one-unit increase in the average diversity index of surrounding neighbourhoods was associated with 7.68% higher property prices. This supports existing literature whereas an optimal mix of employment and housing is associated with higher property prices. For the other relevant models with a lower R^2, the value of the coefficient was even higher. Thus, neighbourhoods depend on the composition of employment sectors in surrounding neighbourhoods.

As mentioned, the regression models that included the number of employees per sector per neighbourhood did not result in a strong association with the average property prices in the neighbourhood with only two significant employment sectors. A first possible reason is that the number of employees within the employment sectors is not sufficiently representative for interpreting the association with the average property price; the number of businesses might be more representative. A second possible reason is that the employment sectors have contradicting values since each employment sector is a merger consisting of a few employment (sub)sectors and might be associated differently with property prices. A third possible reason is that the employment sectors might not have a linear association with property prices, but a rather nonlinear association. A fourth and final possible reason is that people living in neighbourhoods with different population densities have other property preferences and value properties differently. Due to these potential reasons, a stronger association between property prices and employment sectors might have been prevented in the earlier regression models. To account for variations in the regression models, multiple robustness tests were performed.

The first robustness test contained the number of businesses instead of the number of employees per employment sector to factor out the potential impact of large businesses. Considering the three relevant models, coefficients in five out of the eight employment sectors were significantly different from zero. The presence of sector GI (trade and catering industry) was negatively associated with the property prices in the neighbourhood, and partly supporting existing literature. Wholesale is covered in this sector and was expected to be negatively associated with property prices, whereas retail is covered in this sector too and was expected to be positively associated. The presence of sectors HJ (transport, information, and communication), MN (business services), OPQ (government, education, and healthcare), and RSTU (culture, recreation, and other services) was positively associated with the average property prices in the neighbourhood. Sector HJ partly supports existing literature as well; it suggested that information and communication was associated positively, and transport was associated negatively to property prices. On the other hand, sector MN indicated a stronger association with property prices, thereby supporting existing literature. None of the coefficients of the employment sectors were significant in each of the three relevant models, but the association with property prices was considerably stronger compared with the regression models featuring the number of employees per sector.

The second robustness test splits the merged sectors from 8 to 21 employment sectors into single entities, possibly explaining why certain employment sectors in the previous regression models were significantly different from zero. It might be that only one of the employment sectors was significantly different from zero and interfered with another type of employment within the merged sector. The results indicated that sector F (construction) and sector G (wholesale, retail trade, and repair of motor vehicles) were negatively associated with the average property prices in the neighbourhood. The association was strong considering the relevant regression models. Sector F was significant in two models at the 99% confidence interval, whereas sector G was significant in two models as well, although at the 95% confidence interval. Sectors H (transportation and storage), Q (human health and social work activities), and R (culture, sports, and recreation) were positively associated with the average property prices in the neighbourhood. For each of these sectors, only one coefficient was significantly different from zero at the 95% confidence interval out of the three relevant regression models. These positive associations were therefore weaker compared with the associations in sector F and sector G. Considering existing literature, sector F, Q, and R are supportive, whereas sector G is ambiguous. The presence of wholesale was expected to correlate negatively with property prices, whereas retail was expected to positively correlate, both present in sector G. Sector H was expected to negatively correlate with property prices based on earlier literature, but the results indicated the opposite – a negative association.

The third robustness test focused on potential nonlinearity among the employment sectors by squaring all sectors. The results indicated that sector OPQ (government, education, and healthcare) and sector RSTU (culture, recreation, and other services) had coefficients significantly different from zero, thus having a nonlinear relationship. The turning point of the nonlinear curve occurred around the five to eight neighbourhoods with the highest number of employees in these sectors. The nonlinear association was therefore weaker than it appeared initially. A turning point in the U-shaped association for sector OPQ occurred at 1,500 employees, where the average property prices were associated with higher property prices above this number, indicating that residents would only like a high share of governmental, educational, and healthcare workers within the neighbourhood. Sector RSTU exhibited an opposite trend where a turning point in the association arose at 500 employees within the neighbourhood. Up to that point, property values in the neighbourhood were associated with higher property prices, whereafter the property prices were associated with lower property prices. Thus, residents value having some cultural and recreational amenities in proximity, but dislike a great share of this sector, possibly due to additional crowdedness and noise.

The final robustness test was a Chow test to account for differences in the average property prices per neighbourhood for the lowest, middle, and highest population density groups. The null hypothesis – that is, having identical intercepts and slopes in the neighbourhoods with these different population density groups – was rejected, indicating that these subgroups value characteristics within their neighbourhood differently. The results indicated that the neighbourhood within the lowest population density group is associated with higher average property prices when having a greater variety of amenities – that is, the employment sectors – in surrounding neighbourhoods compared with the neighbourhoods with the middle or highest population density groups. Moreover, having a higher number of homes within a neighbourhood was most negatively associated with average property prices in the neighbourhood with the lowest population density group.

To conclude, property prices depend on many different aspects and a specific set of characteristics for each neighbourhood is essential be associated with higher property values. The employment sector variables were applied in multiple variations, all having three relevant regression models. Sector OPQ (government, education, and healthcare) and sector RSTU (culture, recreation, and other services) had significant coefficients in each of the regression variations and were therefore most evidently associated with the average property prices in Amsterdam's neighbourhoods. Both sectors are positively associated with the average property prices; therefore, this study supports existing literature. Sector F (construction) was strongly associated with negative property values and supported existing literature. Sector GI (trade and catering industry) and specifically sector G (wholesale, retail trade, and repair of motor vehicles) were associated with negative average property prices. This only partly supported existing literature since retail was expected to be positively associated with property prices. Having more diverse neighbourhoods in terms of a wide variety of employment sectors and housing is not strongly associated with higher property values. Having more diverse surrounding neighbourhoods, on the other hand, was strongly associated with higher property prices. The neighbourhoods with the lowest population density groups in particular are associated with higher property prices in more diverse surrounding neighbourhoods. This indicates that property prices do not solely depend on the composition within the neighbourhood but even stretch beyond neighbourhood boundaries.

5.2 Policy relevance

The results provide insights into how to organise neighbourhoods within cities, which is valuable information for municipalities, real-estate developers, and investors who are involved in (re)developing neighbourhoods. This knowledge will enable municipalities to stimulate higher property values within the city by focussing on certain employment types within neighbourhoods, and also to focus on diversity within and in surrounding neighbourhoods since this is associated with higher property prices. Repeatedly directing (re)developments in this direction will eventually stimulate higher property values for the entire city. This might attract more people to the city, consequently leading to higher competitiveness compared to other cities. The results are useful for real estate developers as well. By being aware of potential higher property prices in a city and creation of indirect higher property prices, real estate developers might be able to achieve more favourable terms with the municipality. Achieving such terms might directly stimulate higher-quality (re)developments. The final party of interest is the investors, who benefit from higher property values since this will enable them to analyse which (re)developments have the highest potential, either directly within a neighbourhood or indirectly in surrounding neighbourhoods. In particular, if neighbourhoods are to be (re)developed in the future, investors might benefit from long-term investments with greater profits if they get involved early.

5.3 Limitations and future research

This study focused on the association between employment sectors and property prices at the scale of neighbourhoods using Amsterdam as a case study. This study was limited to certain aspects, but simultaneously provides a foundation for future research. The following limitations are proposed along with recommendations for future research:

- In this research, the fourth OLS assumptions was violated; therefore, future research is recommended to explicitly focus on solving independence. Due to the violation, it was not possible to perform an effect study and to interpret a causal relationship between variables; it was only possible to indicate the association between variables and the strength of this association. In the fourth assumption, different employment sectors might attract one another and therefore locate in similar areas, leading to possible endogeneity problems with complex relations. Future research might elaborate further on this issue and solve the independence to interpret a causal relationship between variables.

- This study adopted the scale of neighbourhoods. Because of this, many variables already factored out outliers and presented average values, and therefore, the data were somewhat smoothed. Relating certain characteristics to the average property prices might therefore have been harder than if direct data of property prices and characteristics were available. Future research might focus on smaller clusters, such as postal codes, or use property prices and property characteristics of individual properties. This might result in stronger associations between property prices and the employment sectors.

- This research only focused on the case study of Amsterdam. It is assumed that this type of research is representative of other larger cities in the Netherlands. Future research might indicate whether the results in this research hold for other large cities. Another recommendation is to focus on medium-sized or smaller cities and indicate whether those cities support or contradict the results of this study.

- This study only measured the diversity index of neighbourhoods that contained a mix of work and living functions. It was therefore not possible to investigate the association between adjacent nonresidential neighbourhoods and the average property prices of within neighbourhoods. Future research could provide insights into another type of diversity index to calculate the diversity of nonresidential neighbourhoods or use individual properties and calculate the diversity per property within a certain radius.

- This study used a set of employment variables according to the Dutch SBI (Standard Industrial Classification) and consequently used a merged variation according to CBS (Statistics Netherlands). Some of the employment sectors contradicted existing literature, even in some of the sectors that were already split into single entities. Future research is recommended to further divide the employment sectors, which would make it possible to evaluate the impact of specific businesses on property prices. For example, specifying the retail sector into large, middle and smaller retail shops and account for potential different associations with the property prices.

REFERENCES

Allam, Z., and Jones, D. S. (2020). Pandemic stricken cities on lockdown. Where are our planning and design professionals [now, then and into the future]?. *Land Use Policy*, vol. 97, 104805.

Alvarez, L. (2017). *A place to live, work and play: why mixed-use developments are making a comeback*. [online] The Conversation. Retrieved from: https://theconversation.com/a-place-to-live-work-and-play-why-mixed-use-developments-are-making-a-comeback-73142 [Accessed 20 Jun. 2020].

Buck, R. (2020). *Functiemenging in de stad als instrument – niet als doel.* Vastgoedmarkt. 04 December. [online] Vastgoedmarkt. Retrieved from: https://www.vastgoedmarkt.nl/marktontwikkeling/blog/2020/12/functiemenging-in-de-stad-als-instrument-niet-als-doel-101159268 [Accessed 15 Dec. 2020].

Broekhoven (2017). *De Wederopbouw: Architectuur en Stedenbouw*. [online] Geschiedenis van Zuid-Holland. Retrieved from: https://geschiedenisvanzuidholland.nl/verhalen/de-wederopbouw-architectuur-en-stedenbouw [Accessed 21 May 2021]

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

Cao, T.V. and Cory, D. (1981). Mixed Land uses, Land-use Externalities, and Residential Property Values: a re-evaluation. *Annals of Regional Science,* vol. 16, no. 1, 1-24.

CBS (2018). *Regionale kerncijfers Nederland*. [online] Retrieved from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072ned/table?dl=3B993 [accessed 15 May 2021].

CBS (2019). *Population*. [online] Retrieved from: https://www.cbs.nl/en-gb/background/2018/47/population [accessed 20 Jan. 2021].

CBS (2020). *Toelichting Wijk- en Buurtkaart 2018, 2019 en 2020*. [online] Retrieved from: https://www.cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2020/toelichting-wijk-en-buurtkaart-2018-2019-en-2020?onepage=true [accessed 20 Jan. 2021].

CBS (2021). *Waar groeit of krimpt de bevolking?* [online] Retrieved from: https://www.cbs.nl/nl-nl/dossier/dossier-verstedelijking/hoofdcategorieen/waar-groeit-of-krimpt-de-bevolking- [accessed 20 Jan. 2021].

Commission of European Communities (1990). *Green paper on the urban environment.* Eur 12902. Brussels: EEC.

Coupland, A. (Ed.). (1997). *Reclaiming the city: Mixed use development*. Taylor & Francis.

Couzy, m. and Damen, T. (2018). *Huizenprijzen Amsterdam in een jaar ruim 20 procent gestegen*. [online] Het Parool. Retrieved from: https://www.parool.nl/nieuws/huizenprijzen-amsterdam-in-een-jaar-ruim-20-procent-gestegen~b83e6e2b/ [accessed 27 Apr. 2021]

Daams, M.N., Sijtsma, F.J. and Van der Vlist, A.J. (2016). The Effect of Natural Space on Nearby Property Prices: Accounting for Perceived Attractiveness. *Land Economics,* vol. 92, No. 3, pp. 389-410.

Dantzig, G. B., and T. Saaty (1973). *Compact city: A plan for a livable urban environment*. San Francisco: W.H. Freeman.

De Vor, F., and De Groot, H. L. (2011). The impact of industrial sites on residential property values: A hedonic pricing analysis from the Netherlands. *Regional Studies*, vol. 45, no. 5, 609-623.

Esri Nederland (2021). *Community Map Contributors: “Topo RD”*. Scale not given. [basemap] Retrieved from: http://www.arcgis.com/home/item.html?id=1d1d425122e74d4d87695e163f3f10ce [assessed 03 May. 2021]

Fruth, E., Kvistad, M., Marshall, J., Pfeifer, L., Rau, L., Sagebiel, J., Soto, D., Tarpey, J., Weir, J. and Winiarski, B. (2019). Economic valuation of street-level urban greening: A case study from an evolving mixed-use area in Berlin. *Land Use Policy*, vol. 89, 104237.

Glaeser, E. L., Gyourko, J., and Saks, R. E. (2005). Why have housing prices gone up?. *American Economic Review*, vol. 95, no. 2, pp. 329-333.

Gössling, S. (2020). Why cities need to take road space from cars - and how this could be done. *Journal of Urban Design*, vol. 25, no. 4, pp. 443-448.

Grant, J. (2002). Mixed use in theory and practice: Canadian experience with implementing a planning principle. *Journal of the American Planning Association*, vol. 68, no. 1, pp. 71-84.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*, Random House, London.

Hoppenbrouwer, E., and Louw, E. (2005). Mixed-use development: Theory and practice in Amsterdam's Eastern Docklands. *European Planning Studies*, vol. 13, no. 7, pp. 967-983.

Ismail, S. (2006). Spatial autocorrelation and real estate studies: A literature review. *Malaysian Journal of Real Estate*, vol. 1, no. 1, pp. 1-13.

Jabareen, Y. R. (2006). Sustainable urban forms: Their typologies, models, and concepts. *Journal of Planning Education and Research*, vol. 26, no. 1, pp. 38-52.

Jacobs, J. (2016). *The death and life of great American cities*, Vintage, New York.

Jang, M., and Kang, C. D. (2015). Retail accessibility and proximity effects on housing prices in Seoul, Korea: A retail type and housing submarket approach. *Habitat International*, vol. 49, pp. 516-528.

Koster, H. and Rouwendal, J. (2012). The Impact of Mixed Land Use on Residential Property Values. Vrije Universiteit Amsterdam, *Journal of Regional Science,* vol. 52, no. 5, pp. 733–761.

Koomen, E., Dekkers, J. and Dijk van, T. (2008). Open-space Preservation in the Netherlands: Planning, Practice and Prospects. *Land Use Policy,* vol. 25, pp 361-377.

Lafferty, R.N. and Frech, H.E. (1978). Community Environment and the Market Value of Single Family Homes: the Effect of Dispersion of Land Eses. *Journal of Law and Economics,* vol. 21, pp. 381-394.

Lin, F.J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality and Quantity*, vol. 42, no. 3, pp. 417-426.

Liusman, E., Ho, D. C. W., Lo, H. C., and Lo, D. Y. F. (2017). Office rents, mixed-use developments, and agglomeration economies: a panel data analysis. *Journal of Property Investment & Finance*, vol. 35, no. 5, pp. 455-471.

Midi, H., Sarkar, S. K., and Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, vol. 13, no. 3, pp. 253-267.

Moreno, C., Allam, Z., Chabaud, D., Gall, C. and Pratlong, F., 2021. Introducing the "15-Minute City": Sustainability, resilience and place identity in future post-pandemic cities. Smart Cities, vol. 4, no. 1, pp. 93-111.

MSRC. (2017). *Mixed-Use*. [online] Retrieved from: http://mrsc.org/Home/Explore-Topics/Planning/Development-Types-and-Land-Uses/Mixed-Use.aspx [Accessed 10 Jun. 2020].

Nabil, N. A., and Eldayem, G. E. A. (2015). Influence of mixed land-use on realizing the social capital. *HBRC Journal*, vol. 11, no. 2, pp. 285-298.

Nozeman, E. F. (1977). *De gewenste spatiering van het voorzieningenapparaat, Deelnota D: Instrumenten en kondities gericht op een gespreid resp. gekoncentreerd voorzieningenpatroon in nieuwe woonwijken: een verkennend onderzoek via deskundigen*. Vrije Universiteit Amsterdam.

Palmquist, R. B. (1992). Valuing localized externalities. *Journal of Urban Economics,* vol. 21, pp. 59-68.

Planbureau voor de Leefomgeving (2019). *Sterke groei in steden en randgemeenten verwacht.* [online] Retrieved from: https://www.pbl.nl/nieuws/2019/sterke-groei-in-steden-en-randgemeenten-verwacht [accessed 4 Feb. 2021].

Rouwendal, J., Van Marwijk, R. and Levkovich, O. (2014). The Value of Proximity to Water in Residential Areas. *Real Estate Economics,* vol. 45, no. 4, pp. 829-860.

Rowley, A. (1996). Mixed-use Development: Ambiguous Concept, Simplistic Analysis and Wishful Thinking? *Planning Practice and Research,* vol. 11, no. 1, pp. 85-97.

Rijksoverheid (2021). *Hoe bepalen gemeenten de WOZ-waarde?* [online] Retrieved from: https://www.rijksoverheid.nl/onderwerpen/waardering-onroerende-zaken-woz/vraag-en-antwoord/woz-waarde-bepalen [Accessed 7 Jan. 2021].

Song, Y. and Knaap, G.J. (2004). Measuring the Effects of Mixed Land Uses on Housing Values. *Regional Science and Urban Economics,* vol. 34, pp. 663-680.

Steen, C. (2016). M*ixed-use ontwikkeling, hoe doe je dit?* [online] Gebiedsontwikkeling. Retrieved from: https://www.gebiedsontwikkeling.nu/artikelen/mixed-use-ontwikkeling-hoe-doe-je-het-goed/ [Accessed 20 Jul. 2020].

Urban Hub (2018). *Mixed-use buildings: make the most of your building.* [online] Retrieved from: http://www.urban-hub.com/buildings/mixed-use-buildings-for-diversified-sustainable-sites/ [Accessed 14 Jun. 2020].

Van Duijn, M., and Rouwendal, J. (2013). Cultural heritage and the location choice of Dutch households in a residential sorting model. *Journal of Economic Geography*, vol. 13, no. 3, pp. 473-500.

Verwolf, J. (2019). *De toegevoegde waarde van functiemenging in stadsdeelcentra*. Master thesis, Amsterdam School of Real Estate.

Vorontsova, A. V., Vorontsova, V. L., and Salimgareev, D. V. (2016). The development of urban areas and spaces with the mixed functional use. *Procedia Engineering*, vol. 150, pp. 1996-2000.

Westbeat (2020). Westbeat ontwikkeling. [online] Retrieved from: https://www.westbeat.nl/ [Accessed 28 Jul. 2020].

Wo, J.C. and Kim, J.A. (2020). Neighborhood Effects on Crime in San Francisco: An Examination of Residential, Nonresidential, and ‘Mixed’ Land Uses. *Deviant Behavior*, pp. 1-17.

APPENDICES

Appendix 1 Variables that determine property prices in combination with mixed-use according to existing literature 59

Appendix 2 Maps presenting variables in the Amsterdam neighbourhoods 60

Appendix 3 Summary of all used variables 62

Appendix 4 Model adjustments...65

Appendix 5 Testing the OLS-assumptions..69

Appendix 6 Log transformation of WOZ-value 77

Appendix 7 Descriptive statistics 78

Appendix 8 Descriptive statistics of other variants of the employment sector variables (numbers and share) 80

Appendix 9 Correlation between the employment sectors and the average property price per neighbourhood 82

Appendix 10 Regression models (complete table)..84

Appendix 11 Elaboration of the remaining significant variables in Appendix 10 87

Appendix 12 Nonlinear figures...91

Appendix 13 Calculation Chow test 92

Appendix 14 Regression models Chow test 93

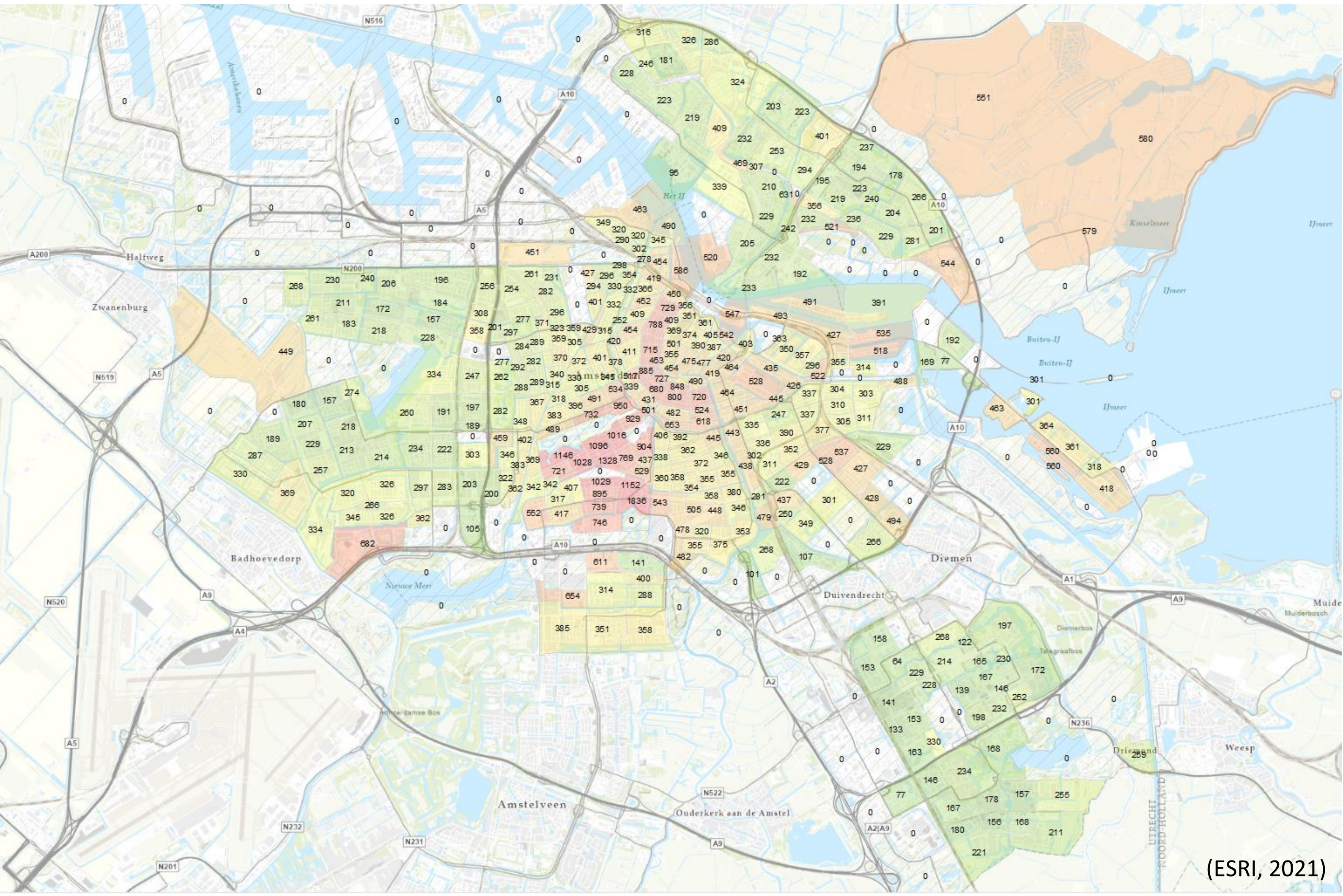
Appendix 15 Stata syntax...96

Appendix 1 Variables that determine property prices in combination with mixed-use according to existing literature

Variables	***Authors***
Employment sectors in neighbourhood	
- Commercial (%)	Cao and Cory (1982); Lafferty and Frech (1978)
- Industrial (%)	Cao and Cory (1982); Lafferty and Frech (1978)
- Public or institutional (%)	Cao and Cory (1982); Lafferty and Frech (1978)
- Vacant land (%)	Cao and Cory (1982); Lafferty and Frech (1978)
- Agricultural (%)	Lafferty and Frech (1978)
- Distance to and share in neighbourhood of:	
o Commercial (km + %)	Song and Knaap (2004)
o Public/institutional (km + %)	Song and Knaap (2004)
o Industrial (km + %)	Song and Knaap (2004)
Housing characteristics	
- Lot size (m2)	Song and Knaap (2004)
- Floor area building (m2)	Cao and Cory (1982); Koster and Rouwendal (2012); Lafferty and Frech (1978); Song and Knaap (2004)
- Volume (m3)	Koster and Rouwendal, (2012)
- Construction year/building age (year)	Cao and Cory (1982); Koster and Rouwendal (2012); Lafferty and Frech (1978); Song and Knaap (2004)
- Rooms (#)	Cao and Cory (1982); Koster and Rouwendal (2012); Lafferty and Frech (1978)
- Garage (y/n)	Koster and Rouwendal (2012)
- Basement (y/n)	Cao and Cory (1982)
- Garden (y/n)	Koster and Rouwendal (2012)
- No central heating (y/n)	Koster and Rouwendal (2012)
- Listed building (y/n)	Koster and Rouwendal (2012)
- Number of homes (#)	Cao and Cory (1982); Lafferty and Frech (1978)
o Apartments (%)	Koster and Rouwendal (2012)
o Terraced (%)	Koster and Rouwendal (2012)
o Semi-detached (%)	Koster and Rouwendal (2012)
o Detached (%)	Koster and Rouwendal (2012)
Spatial characteristics	
- Distance to center (km)	Koster and Rouwendal (2012)
- Distance to CBD (ft.)	Cao and Cory (1982); Song and Knaap (2004)
- Distance to train station (km)	Koster and Rouwendal (2012)
- Distance to open space (km)	Koster and Rouwendal (2012)
- Distance to main road (ft.)	Song and Knaap (2004)
- Distance to minor road (ft.)	Song and Knaap (2004)
- Proximity to bus stop (y/n)	Song and Knaap (2004)
- Public park (km + %)	Song and Knaap (2004)
- Proximity to water bodies (y/n)	Song and Knaap (2004)
Neighbourhood characteristics	
- Income ($)	Song and Knaap (2004)
- Household density (people/acre)	Song and Knaap (2004)
- Proportion multifamily (%)	Song and Knaap (2004)
- Ethnic minority (%)	Cao and Cory (1982); Koster and Rouwendal (2012); Song and Knaap (2004)
- (Un)employment status (y/n)	Cao and Cory (1982)
- Proportion poor families (%)	Cao and Cory (1982)
- Crowdedness (%)	Cao and Cory (1982)
- Noise (y/n)	Cao and Cory (1982)
- Crime (#/1000 people)	Cao and Cory (1982)

Appendix 2 Maps presenting variables in the Amsterdam neighbourhoods

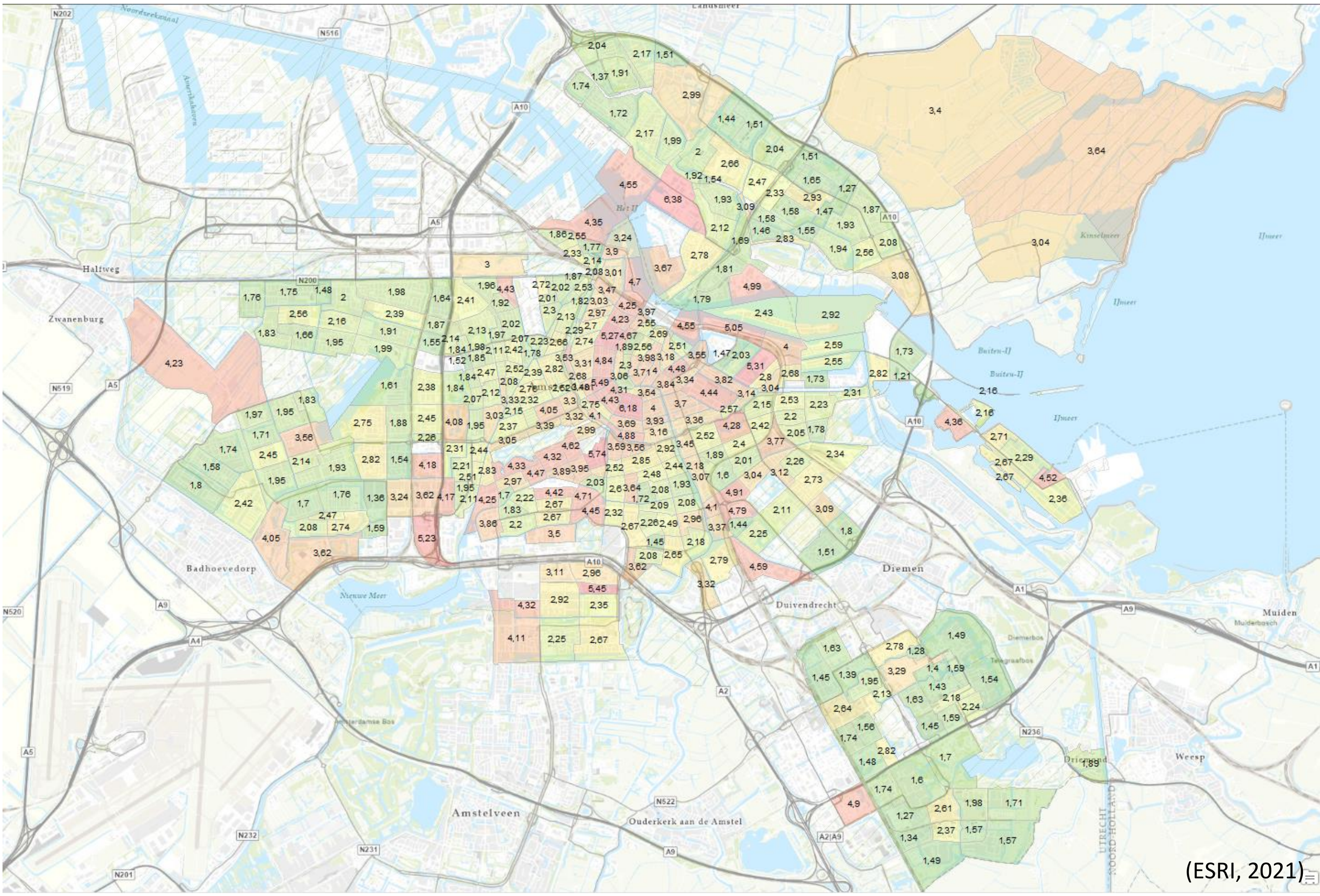
Appendix 2a: The average WOZ value in Amsterdam per neighbourhood in 2018



(ESRI, 2021)

Note: This figure indicates a map of Amsterdam and visualizes the average property price per neighbourhood. Green coloured neighbourhoods indicate a relatively low average property value whereas red coloured neighbourhoods indicate a relatively high average property value. The neighbourhoods with faded grey lines are the nonresidential neighbourhoods and are not included in this research. The average property value per neighbourhood is based on the CBS dataset as mentioned in Section 3.1.

Appendix 2b: The diversity index in Amsterdam per neighbourhood in 2018



(ESRI, 2021)

Note: This figure indicates a map of Amsterdam and visualizes the average diversity index per neighbourhood. Green coloured neighbourhoods indicate a relatively low diversity index whereas red coloured neighbourhoods indicate a relatively high diversity index. The neighbourhoods with faded grey lines are the nonresidential neighbourhoods and are not included in this research. The average diversity index is calculated as mentioned in Section 3.2.

Appendix 3 Summary of all used variables

Variables	***Variable description***
Main variables	
ln_woz	The log is taken of the average value of immovable property (or WOZ value) in Euros
di	The diversity index of the neighbourhood
avg_di	The average diversity index of surrounding neighbourhoods
w_avg_di	The weighted average of the diversity index, whereas the weight is applied to the number of jobs in the surrounding neighbourhoods
wk_code	The greater neighbourhood scale, consisting of 96 neighbourhoods.
Employment sectors in neighbourhood	
Employment sectors (merged):	
○ Employees sector A	A: Agriculture, forestry and fishing (# of employees)
○ Employees sector BCDEF	B, C, D, E & F: Industry and energy (# of employees)
○ Employees sector GI	G + I: Wholesale, retail and catering industry (# of employees)
○ Employees sector HJ	H + J: Transport, information and communication (# of employees)
○ Employees sector KL	K + L: Financial services and real estate (# of employees)
○ Employees sector MN	M + N: Business services (# of employees)
○ Employees sector OPQ	O, P & Q: Government, education, healthcare (# of employees)
○ Employees sector RSTU	R, S, T & U: Culture, recreation, other services (# of employees)
Employment sectors (as single entities):	
○ Employees sector A	A Agriculture, forestry and fishing (# of employees)
○ Employees sector B	B Mining and quarrying (# of employees)
○ Employees sector C	C Manufacturing (# of employees)
○ Employees sector D	D Electricity, gas, steam and air conditioning supply (# of employees)
○ Employees sector E	E Water supply; sewerage, waste management and remediation activities (# of employees)
○ Employees sector F	F Construction (# of employees)
○ Employees sector G	G Wholesale and retail trade; repair of motor vehicles (# of employees)
○ Employees sector H	H Transportation and storage (# of employees)
○ Employees sector I	I Accommodation and meal and drink service activities (# of employees)
○ Employees sector J	J Information and communication (# of employees)
○ Employees sector K	K Financial institutions (# of employees)
○ Employees sector L	L Real estate rental and trade (# of employees)
○ Employees sector M	M Consultancy, research and other specialised business services (# of employees)
○ Employees sector N	N Renting and leasing of tangible goods and other business support services (# of employees)
○ Employees sector O	O Public administration, government services and compulsory social security (# of employees)
○ Employees sector P	P Education (# of employees)

○ Employees sector Q	Q Human health and social work activities (# of employees)
○ Employees sector R	R Culture, sports and recreation (# of employees)
○ Employees sector S	S Other service activities (# of employees)
○ Employees sector T	T Households as employers; undifferentiated goods and service-producing activities of households for own use (# of employees)
○ Employees sector U	U Extraterritorial organisations and bodies (# of employees)
Structural characteristics in the neighbourhood	
- Housing characteristics	
○ Homes	Number of homes (#)
○ Floor space	
▪ c_0_50	Floor space between 0-50 sq.m. (%)
▪ c_51_75	Floor space between 51-75 sq. m. (%)
▪ c_76_100	Floor space between 51-100 sq.m. (%)
▪ c_101_150	Floor space between 101-150 sq.m. (%)
▪ c_151_250	Floor space between 151-250 sq.m. (%)
▪ c_251_groter	Floor space between ≥ 251 sq.m. (%)
○ Construction period	
▪ c_pre1900	Construction year before 1900 (%)
▪ c_1900_1944	Construction year 1900-1944 (%)
▪ c_1945_1979	Construction year 1945-1979 (%)
▪ c_1980_1999	Construction year 1980-1999 (%)
▪ c_post2000	Construction year after 2000 (%)
▪ c_unknown	Construction year unknown (%)
- Property types	
○ p_1gezw	Single-family housing (%)
○ p_mgzw	Multifamily housing (%)
○ p_bewndw	Inhabited housing (%)
○ p_leegsw	Vacant housing (%)
○ p_koopwon	Owner-occupied homes (%)
○ p_huurwon	Rental homes (%)
Spatial characteristics in neighbourhood	
- Land surface	
○ p_opp_land	Area of land (%)
○ p_opp_water	Area of water (%)
- Average distance (in km) per household to nearest:	
○ af_oprit	Main road driveway
○ af_treinst	Train station
○ af_overst	Important transfer station
○ af_artspr	General practice
○ af_artspo	General practice station
○ af_apoth	Pharmacy
○ af_ziek_i	Hospital including outpatient clinic
○ af_ziek_e	Hospital excluding outpatient clinic
○ af_superm	Supermarket
○ af_daglmd	Other daily food shops
○ af_warenh	Department store
○ af_cafe	Cafe
○ af_caftar	Cafeteria
○ af_restau	Restaurant
○ af_hotel	web
○ af_kdv	Day-care centre
○ af_bso	Out-of-school care

○ af_ondbas	Primary school
○ af_ondrvrt	Secondary education
○ af_ondvmb	Vmbo schools
○ af_ondhv	Havo/vwo schools
○ af_brandw	Fire station
○ af_zwemb	Swimming pool
○ af_ijsbaan	Ice rink
○ af_biblio	Library
○ af_pop	Music venue
○ af_bios	Cinema
○ af_sauna	Sauna
○ af_zonbnk	Solarium
○ af_attrac	Attraction park (amusements park, zoo or indoor playground)
○ af_podium	Performing arts
○ af_museum	Museum
Social environment in neighbourhood	
- Residential characteristics	
○ bev_dichth	Population density (inhabitants/sq.km)
○ oad	Average address density (/sq.km)
○ sted	Urbanity level (code 1-5) - 1 = extremely urban (≥ 2500 addresses/ sq. km) - 2 = very urban (1500-2500 addresses/ sq. km) - 3 = moderately urban (1000-1500 addresses/ sq. km) - 4 = not very urban (500-1000 addresses/ sq. km) - 5 = non-urban (<500 addresses/ sq. km)
- Immigrants:	
○ p_west_al	Western immigrants (%)
○ p_n_w_al	Non-Western immigrants (%)
- Income:	
○ p_laaginkh	Households having the lowest income (%)
○ p_hooginkh	Households having the highest income (%)
○ p_sociminh	Households at or below the social minimum (%)
○ p_nietact	Non-active residents between 15-75 years (%)
- Social security benefits	
○ wwb_uittot	General social assistance benefits (#)
○ a_soz_ow	Total AOW (General Old Age Pensions Act) benefit (#)
○ ao_uit_tot	Total AO (General Disability Act) benefits (#)
○ ww_uit_tot	Total unemployment benefits (#)
- Criminality	
○ g_wodief	Theft (# per 1000 residents)
○ g_vernoo	Destruction, crime against public order (# per 1000 residents)
○ g_gewsek	Violent and sexual crimes (# per 1000 residents)

Appendix 4 Model adjustments

Within this appendix, all model adjustments are described by showing the dummy variables and the adjustments for the multicollinear variables. The dataset only consists of average values of the neighbourhoods. These values originate from all individual observations within those neighbourhoods due to which outliers are not present within the dataset. Two variables are excluded from the model since having too many values containing value zero. These are the variables with the number of income receivers and the average income per resident within a neighbourhood, where 295 and 243 variables out of 382 variables are missing, respectively.

Dummies

In order to create valid regression models, a few variables are interpreted as dummy variables and are mutually exclusive. Meaning for example that houses are either rentals or owner-occupied, indicated by a 0 and 1. The first two models do not contain dummy variables. The third model includes dummy variables for the employment sectors. The agricultural sector (*p_wrkn_a*) and the financial services and real sector (*p_wrkn_kl*) are less suited to serve as reference category since 330 and 30 neighbourhoods do not have any employees in these sectors, respectively. The remain sectors are suitable and do not have more than one neighbourhood without employees. The chosen reference is the industry sector (*p_wrkn_bcdef)*, a variable with no missing values.

The fourth model includes dummy variables for the floor space in square metres, divided in six categories: <50, 51-75, 76-100, 101-150, 151-250 and 251< (the variables c_0_50, c_51_75, c_76_100, c_101_150, c_151_250 and c_251_grot). The reference category is the variable smaller than 50 square metres and therefore not included in the actual regression model. To interpret the variables with a larger floor space, it references to the floor space smaller than 50 square metres. For example, interpreting the coefficient of 51-75 square metres which is 0.452. It indicates that the association on the average property price in the neighbourhood is 57% higher (($e^{0.45}$ – 1) *100%) than the average housing prices in the neighbourhood with floor spaces smaller than 50 square metres. The building year for each building is also presented by dummy variables, and merged into six categories: before 1900, 1901-1944, 1945-1979, 1980-1999, post 2000 and unknown (the variables c_pre1900, c_1900_1944, c_1945_1979, c_1980_1999, c_post2000 and c_unknown). The reference category is post 2000. Other dummy variables are the percentage multifamily homes (0) and single-family homes (1) *(p_mgezw* and *p_1gezw)*; rental homes (0) and owner-occupied homes (1) *(p_huurwon* and *p_koopwon)*, and; occupied homes (0) and vacant homes (1) *(p_bewndw* and *p_leegsw)*. All these dummy variables are also used in model 5 and 6.

The fifth model only contains a dummy variable for the percentage land (0) and the percentage water (1) *(p_opp_land* and *p_opp_water)*. The other variables do not contain dummy variables since all variables indicate the average distance to amenities. The sixth model includes dummy variables for multi-person households (0) and single-family households (1) *(p_eenp_hh* and *p_mgz_hh)*.

Multicollinearity

Another factor is to check the absence of multicollinearity, in other words, that independent variables are not correlated with each other (Brooks and Tsolacos, 2010). By having perfect multicollinearity, two or more variables have an exact relationship, resulting in a correlation 1 or -1. Otherwise, a non-negligible, but not perfect relationship between two or more variables is near-multicollinearity. The R2, or explained variance, will then be high and the standard error of the individual coefficients will be high as well, preventing significant results in the individual variables. An option to deal with multicollinearity is ignoring it since near-multicollinearity does not violate the OLS assumption and does not affect the BLUE properties of OLS. Another option is to drop one of the collinear variables or transform highly correlated variables into ratio variables and exclude the original variable.

Each model is checked for multicollinearity by looking at its correlation matrix and its VIF (Variance Inflation Factor) by the use Stata. A rule of thumb is that correlations between the independent variables with a value higher than 0.8 are considered collinear (Midi, Sarkar and Rana, 2010). The correlation matrix only accounts for correlations pairwise, whereas the VIF also checks multicollinearity for two or more variables. When the VIF shows a value higher than 10, it indicates multicollinearity, otherwise variables are not multicollinear (Lin, 2008). The following paragraphs indicate the adjustments in the models.

The first two models do not contain any multicollinearity. Within the third model, the variable that contains the total number of employees (*a_wrkn_tot*) in each neighbourhood is multicollinear with almost all other employment sectors and shows perfect correlation with sector BCDEF (*a_sect_bcdef*). The reason of multicollinearity is that the total number of employees is the exact sum of all separate sectors what therefore creates multicollinearity. By excluding the total number of employees, the problem of multicollinearity is solved. Since the models are extended in each successive model, the total number of employees in each neighbourhood is excluded in model 4, 5 and 6 as well.

The fourth model does have one multicollinear variable, the average building year (*avg_bwj*). This variable is excluded from the model. The fifth model with spatial characteristics does have multicollinearity, 13 variables have a VIF ranging between 10 and 30. An option is to keep all variables since the model still is BLUE and therefore the coefficients are not biased. Despite, it cannot be concluded which coefficients are significantly associated with the average property prices. Therefore, the (multi)collinear variables are excluded from the model. The variables being highly collinear with a value over 0.9 in the correlation matrix are: distance to general practice post (*af_artspo*) and distance to hospital excluding outpatient clinic (*af_ziek_e*). From these variables, the variable *af_arts_po* is excluded from the model since it is the one with the highest multicollinearity. Other variables having that are multicollinear according to its VIF are distance to: music venue (*af_pop*), cinema (*af_bios*), performing arts (*af_podium*), other daily food shops (*af_dagmld*) and cafeteria (*af_caftar*). Other variables that still are collinear by having a value higher than 0.8 according to its correlation matrix are distance to: af_apoth and af_artspr; af_apoth and af_superm; af_bso and af_kdv, and; af_museum and af_sauna. The variables af_apoth, af_bso and af_sauna are excluded from the model. The remaining variables are kept and used within the sixth model.

The sixth model causes multicollinearity in the variable of distance to museum (*af_museum*). With a value of 9.55, this variable had the highest VIF of within model 5 and now exceeds the limit of being multicollinear with a value of 11.47. Therefore, this variable is excluded from model 6. Multicollinearity is also found between the variable of residents having the lowest income (*p_laaginkp*) and households having the lowest income (*p_laaginkh*). Since the diversity index is calculated by the number of households, the variable containing the number of households is maintained and the variable containing residents is excluded from the variable list. The same holds for residents having the highest income (*p_hooginkp*) and households having the highest income (*p_hooginkh*), the variable containing residents with the highest income is excluded. The share of high- and low-income households show multicollinearity with the variables single households (*p_eenp_hh*) and average household size (*gem_hh_gr*). Song and Knaap (2004) use the variable income and Cao and Cory (1982) use the variable proportion of poor families. For this reason, prosperity of households better corresponds with earlier literature and the share of single households and average household size are excluded from the model.

Furthermore, the variable with the number of homes (*homes*), the number of households (*aant_hh*) and the number of residents (*aant_inw*) are multicollinear. To prevent that, only one variable is kept in the model. The number of homes is considered the most important variable since this is referring to the physical characteristics of the neighbourhood. The number of households and number of residents are excluded from the model and considered less important since they are referring to the social environment in the neighbourhood. To mention still something about the population, the average population density (*bev_dichth*) can be consulted. The variables with a value of higher than 0.8 in the correlation matrix are the number of homes (homes) and the number of unemployment benefits (*ww_uit_tot*); the number of unemployment benefits (*ww_uit_tot*) and the number of AO, or General Disability Act, benefits (*ao_uit_tot*), and; the number of AO, or General Disability Act, benefits (*ao_uit_tot*) and the number of general social assistance benefits (*wwb_uittot*). Since the number of homes is the most important variable, this variable is kept, and the number of unemployment benefits is excluded. Since Cao and Cory (1982) use the variable of poor people, the variable general assistance benefits overlap the most with existing literature. For this reason, the number of people in the General Disability Act are excluded from the model.

Appendix 5 Testing the OLS-assumptions

Within this appendix, each of the OLS-assumptions are tested by using Stata. When the first four assumptions hold, the coefficients to be estimated have a number of desirable properties, known as a best linear unbiased estimator (BLUE) according to Brooks and Tsolacos (2010). Specifically, "estimator" indicates that the estimators represent the true value of the actual coefficients. "Linear" stands for linearity in the estimators, where the formula for the estimators are linear combinations of the dependent variable. "Unbiased" means that the actual values of estimators are on average equal to their true values. Finally, "best" represents a minimal variance in the OLS estimator among the class of linear unbiased estimators. In other words, the estimators have desirable properties to be consistent, unbiased and efficient.

Assumption 1: $E(u_t) = 0$

In the first assumption, it must be made sure that the average value of the errors is zero. Since an intercept is included in the regression model, this assumption will never be violated and ensures linearity in the model according to Brooks and Tsolacos (2010).

Assumption 2: $var(u_t) = \sigma^2 < \infty$

In the second assumption, it is checked whether the variance of the errors is constant, in other words, if the models are homoscedastic. This is tested by a graphical and non-graphical method. The graphical method is tested by rvfplot, plotting the residuals versus the fitted values. A well-fitted model is related to no pattern between the residuals and the fitted values, and thus being homoscedastic, otherwise the model is heteroscedastic. The non-graphical method is using the Breusch-Pagan test with the null hypothesis of having a constant variance. The test is performed for each of the regression models in Stata and presented in Appendix table 5.1 at the end of this Appendix. It demonstrates that all models are significant, meaning that the null hypothesis of having a constant variance is rejected. This indicates that the models are not homoscedastic but heteroscedastic. The models still have consistent and unbiased coefficient estimates, but are no longer BLUE, or specifically referring to no longer have a minimum variance among the class of unbiased estimators (Brooks and Tsolacos, 2010). As the authors further mention, a solution to deal with that is to use heteroscedasticity-consistent standard error estimates, also called robust standard errors. By using these, the standard error estimates have been modified to account for heteroscedasticity. All models are therefore conducted with robust standard errors.

Assumption 3: $cov(u_i, u_j) = 0$

The third assumption supposes that the covariance between the error terms over time is zero, therefore having uncorrelated error terms. This research conducts a cross-sectional approach and is not measured over time. According to Brooks and Tsolacos (2010), covariance might still occur between error terms in spatial terms. By having errors that are not uncorrelated, the errors are autocorrelated. As the authors mention, the models are no longer BLUE (best linear unbiased estimator), indicating that the coefficients are no longer consistent and the standard errors no longer efficient, similar as in the previous assumption. In real estate, it often occurs that observations adjacent to each other have similar characteristics, also called “locational similarity” (Ismail, 2006). It is therefore useful to run a regression in clustered groups based on location to account for the locational similarity. This clusters the standard errors and therefore deals with autocorrelation in spatial terms. Hence, this research is already based on clusters and not on individual observations. The clusters are the neighbourhoods, whereas the values of the neighbourhoods are based on individual observations within the neighbourhoods. The current level of neighbourhoods consists of 382 neighbourhoods (in Dutch: “buurten”), the clusters are applied on a greater neighbourhood scale (in Dutch: “wijken”), consisting of 96 clusters. Each regression model is performed by using clustered standard errors on this greater neighbourhood scale. As a result, the assumption of not having autocorrelation is not violated. The function of using clustering standard errors in Stata simultaneously accounts for the robust standard errors as mentioned in assumption 2.

Assumption 4: $cov(u_t, x_t) = 0$

In the fourth assumption, by having no relationship between the error and the independent variable, it is also referred to as having no endogeneity. Endogeneity may occur if: 1) correlated variables are missing, 2) the selected sample is correlated with the error term, or 3) there is reverse causality, where the dependent variable also causes variation in the independent variable (Brooks and Tsolacos, 2010). For the models in this research, it seems that no correlated variables are missing since the used variables greatly correspond with the variables found in earlier literature. In addition, the sample is right since no neighbourhoods in Amsterdam are excluded and all neighbourhoods containing WOZ values are used. The latter reason of endogeneity is reverse causality, which could possibly be between the housing price and income. Residents having a high income are likely to buy larger homes, but large homes might at the same time attract people with a high income. Despite the variable income is excluded from the model due to too many observations having a value of zero as described in Appendix 4.

Appendix table 5.2: The residuals

VARIABLES	(4) + Housing charact.	(5) + Spatial charact.	(6) + Social environment
Res_lnwoz_m4	-0.769*** (0.149)		
Res_lnwoz_m5		-0.306** (0.154)	
Res_lnwoz_m6			-0.115** (0.0519)
Observations	382	382	382
R-squared	0.824	0.898	0.951

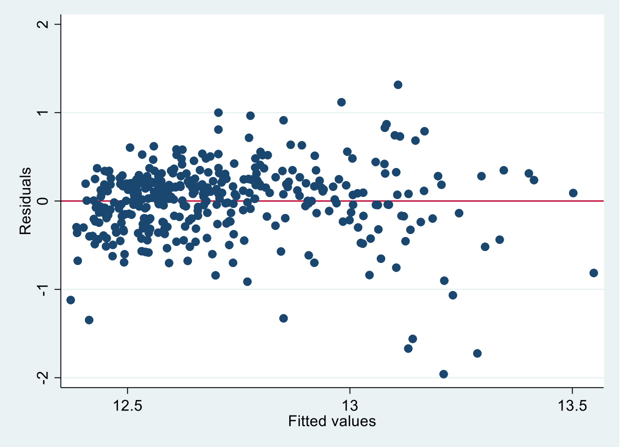
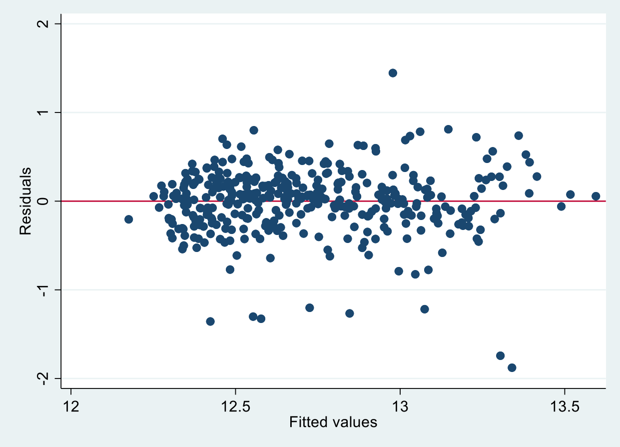
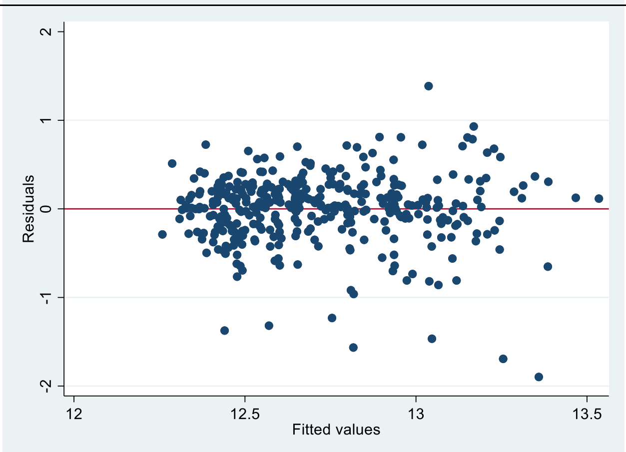
To test for endogeneity, the Durbin-Wu-Hausman test is applied, having a null hypothesis that there is independence between the error term and the dependent variable. First, the residuals are calculated for each of the relevant regression models without the log of the average property price per neighbourhood, therefore these regressions are performed on the diversity index per neighbourhood. Then, the regression is performed by adding the residuals to regression model that includes the log of the average price per neighbourhood again. The coefficients of the residual are presented in Appendix table 5.2 and indicate that all coefficients are significant, thus the null hypothesis is rejected in all models. This indicates the presence of endogeneity. As a result of that, this study does not serve as an impact study, meaning that effects or causal relationships between variables cannot be determined. Therefore, it is only possible for this study to measure the correlation between variables, thus indicating the association between variables.

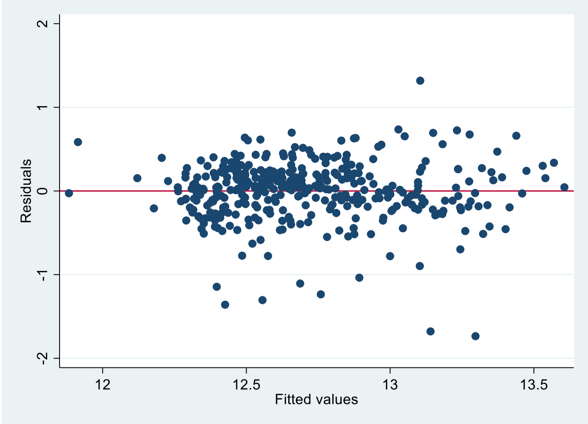
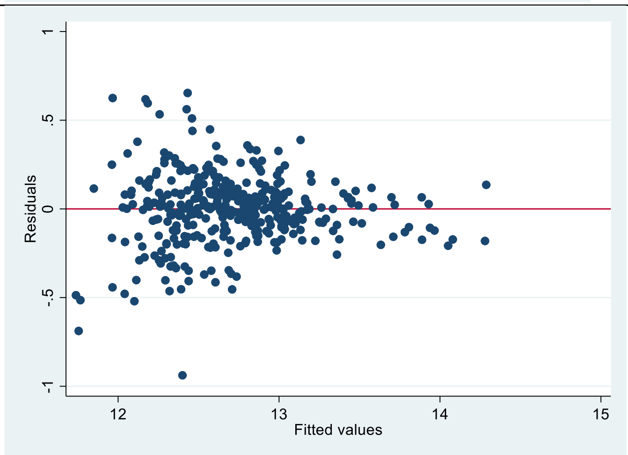
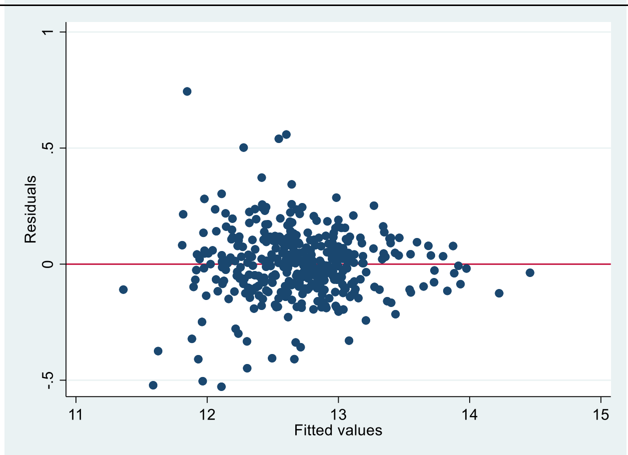
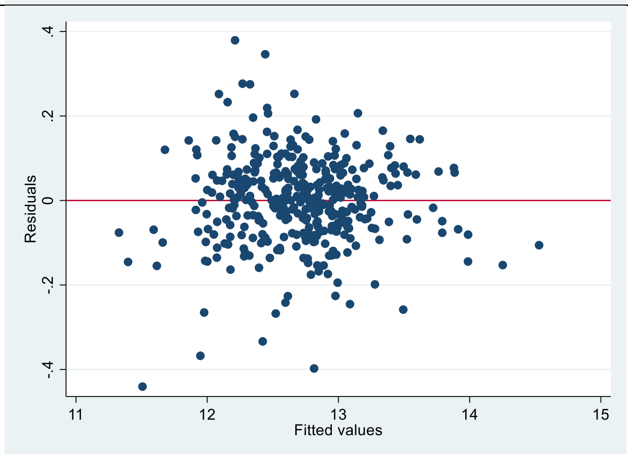
In addition to that, the study of Koster and Rouwendal (2012) also argues that measures of diversity between either employment sectors, or housing and specific employment sectors might be endogenous as well. Different employment sectors might locate in areas with a similar diversity index and therefore value land similarly in a bid-rent context. Locations might be attractive for residents as well as firms, but the reasoning for that might be unobserved. For example, that some types of shops attract certain types of residents, and vice versa, leading to reverse causality as well. Therefore, Koster and Rouwendal (2012) leave the potential endogeneity issues out of their scope and recommend this for further investigation. Since the diversity index is based on different employment sectors and housing, examining correlations between employment sectors might lead to complex relations. For this reason, the endogeneity issue will be left out of consideration.

Assumption 5: $u_t \sim N(0, \sigma^2)$

In order to check whether there is normality among the residuals, each model is tested in Stata. Each regression model is run, whereafter a prediction command is used to create residuals. These residuals are plotted in a graph by using the kernel density plot. By also adding the normal density, it shows whether there is normality among the residuals. Each model is verified by the Shapiro-Wilk W test, a numerical test for normality. The results from Stata are presented in Appendix table 5.3, at the end of this appendix. All regression models are significant, indicating that the null hypothesis of having a normal distribution of the residuals is rejected. According to Brooks and Tsolacos (2010), an important result in statistics is the Central Limit Theorem. If the sample size increases, the sampling distribution will approximately converge to the normal distribution, even if the population distribution itself is not normal. As the authors further mention, if the sample size is not sufficiently large, it involves the risk of making invalid inferences. To invoke the Central Limit Theorem, having more than 100 observations is sufficiently large. This study uses 382 neighbourhoods, well above the critical value of a sufficiently large sample size, and therefore normality is assumed.

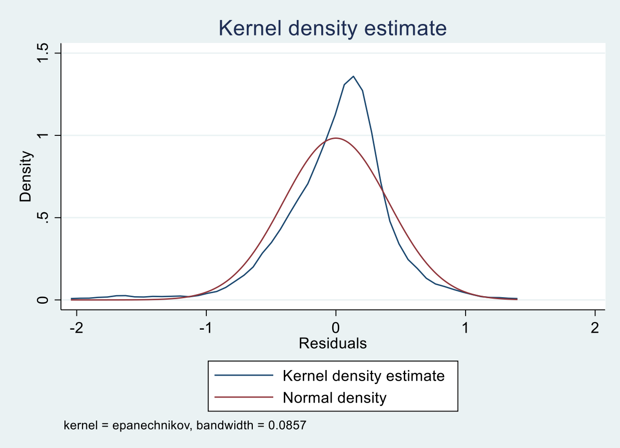
Appendix table 5.1: Testing heteroscedasticity by a rvfplot and the Breusch-Pagan test

Model	Residuals versus fitted values plot	Breusch-Pagan test
1		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **81.52** Prob > chi2 = **0.0000**
2a		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **30.51** Prob > chi2 = **0.0000**
2b		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **47.99** Prob > chi2 = **0.0000**

3		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **17.76** Prob > chi2 = **0.0000**
4		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **70.41** Prob > chi2 = **0.0000**
5		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **60.39** Prob > chi2 = **0.0000**
6		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of ln_woz chi2(1) = **20.88** Prob > chi2 = **0.0000**

Appendix table 5.3: Testing normality by the Kernel density plot and the Shapiro-Wilk W test

<table>
<tr><th>Model</th><th>Kernel density plot</th><th>Shapiro-Wilk W test</th></tr>
<tr><td>1</td><td>

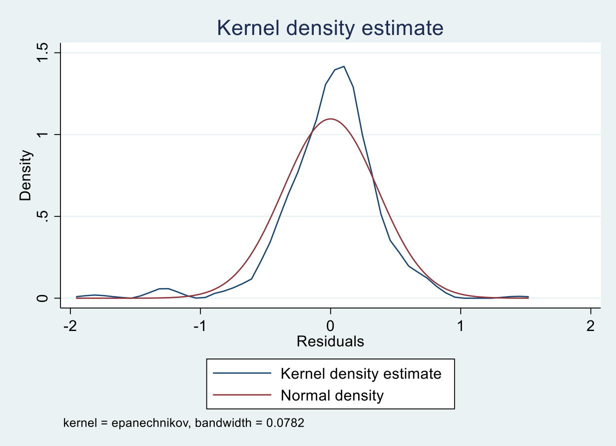


</td><td>

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_model1	382	0.94071	15.701	6.539	0.00000

</td></tr>
<tr><td>2a</td><td>

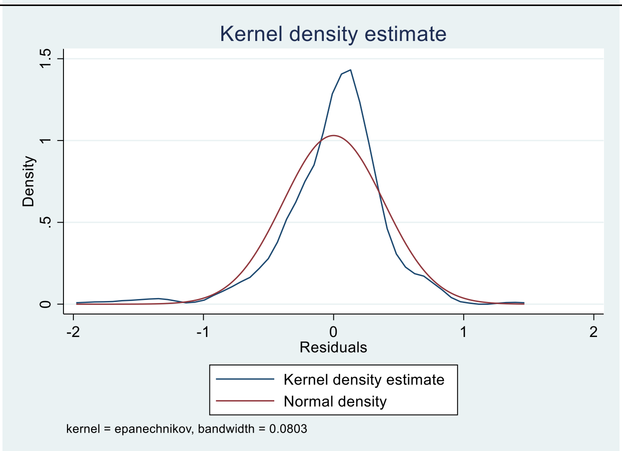


</td><td>

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_model2a	382	0.92591	19.574	7.062	0.00000

</td></tr>
<tr><td>2b</td><td>



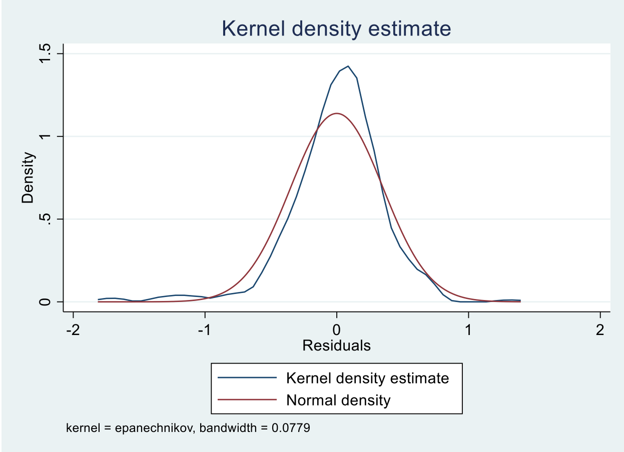
</td><td>

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_model2b	382	0.93273	17.772	6.833	0.00000

</td></tr>
</table>

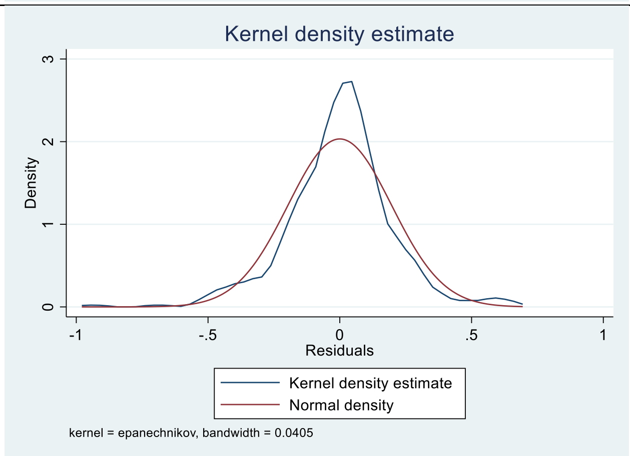
3



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_model3	**382**	**0.93254**	**17.823**	**6.839**	**0.00000**

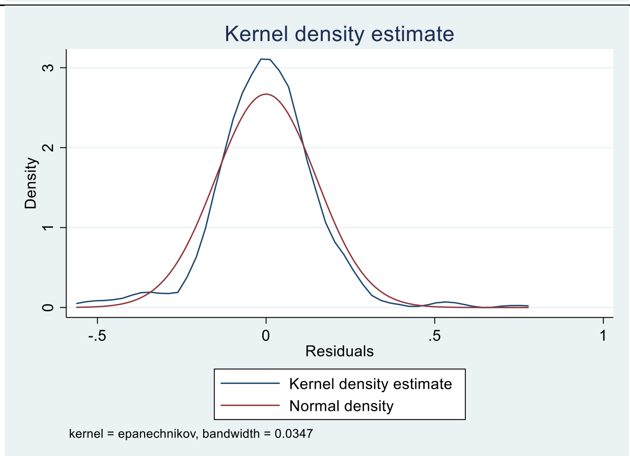
4



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_model4	**382**	**0.96754**	**8.574**	**5.102**	**0.00000**

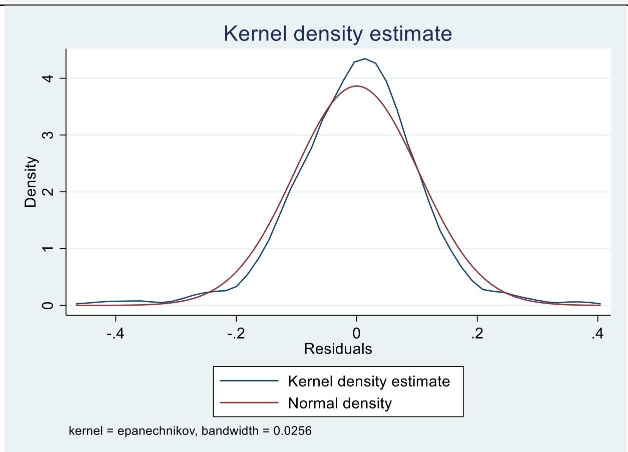
5



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_model5	**382**	**0.95645**	**11.505**	**5.800**	**0.00000**

6

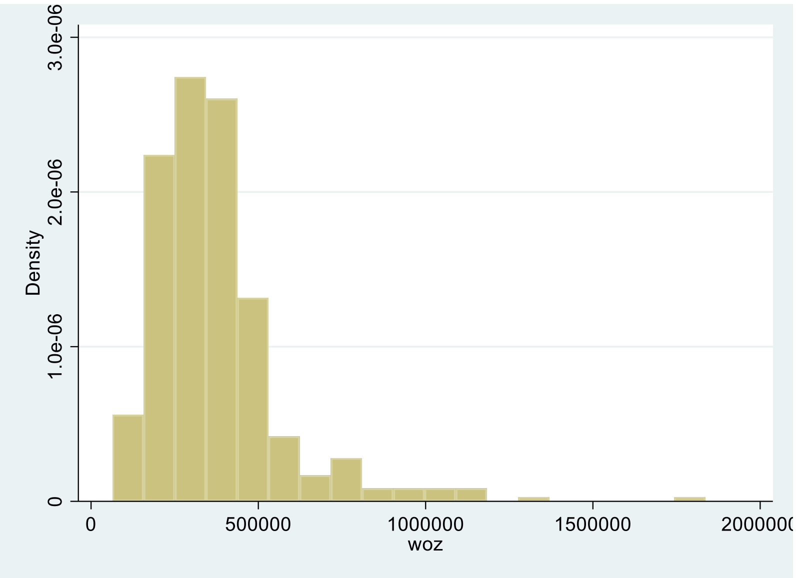


Shapiro-Wilk W test for normal data

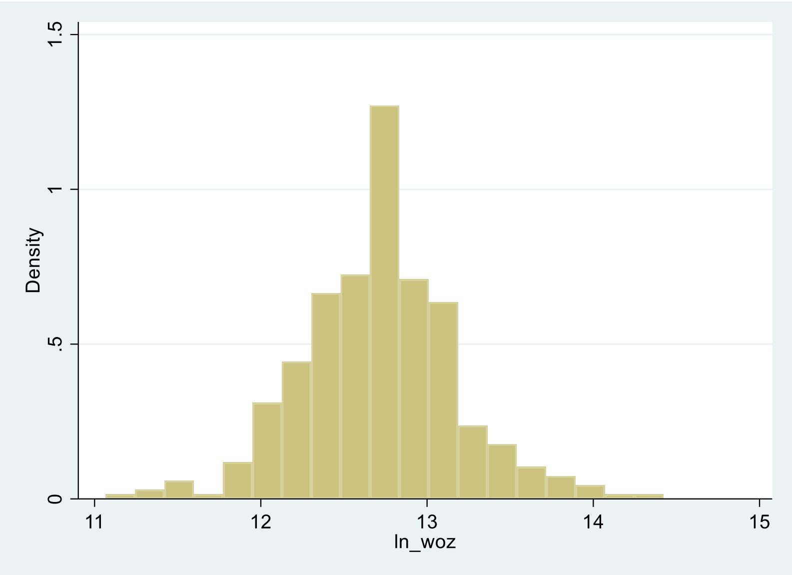
Variable	Obs	W	V	z	Prob>z
res_model6	**382**	**0.97443**	**6.756**	**4.536**	**0.00000**

Appendix 6 Log transformation of WOZ-value

The following figures represent the log transformation of the dependent variable, the average WOZ value per neighbourhood.



Appendix figure 6.1: Histogram of the average WOZ value per neighbourhood



Appendix figure 6.2: Histogram of the log transformed average WOZ value per neighbourhood

Appendix 7 Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
log of average woz value	382	12.709	.468	11.067	14.423
diversity index	382	2.691	1.024	1.212	6.383
average diversity index	382	2.719	.674	1.422	4.550
weighted average diversity index	382	2.922	.753	1.431	5.106
employees sector a (x100)	382	.005	.025	0	.31
employees sector bcdef (x100)	382	.524	1.25	0	16.45
employees sector gi (x100)	382	2.849	3.895	0	30.71
employees sector hj (x100)	382	1.11	1.582	0	13.09
employees sector kl (x100)	382	1.052	6.437	0	96.78
employees sector mn (x100)	382	2.692	4.007	.02	48.26
employees sector opq (x100)	382	2.758	4.524	.01	53.66
employees sector rstu (x100)	382	1.161	1.137	.01	9.92
homes	382	1123.937	668.556	53	3359
avg opp	382	81.851	28.491	23	346
c 0 50	382	.219	.194	0	1
c 51 75	382	.336	.188	0	.83
c 76 100	382	.243	.162	0	.87
c 101 150	382	.148	.139	0	.8
c 151 250	382	.043	.072	0	.48
c 251 grot	382	.011	.032	0	.41
avg bwj	382	1950.042	40.582	1780	2018
c pre1900	382	.081	.187	0	.96
c 1900 1944	382	.299	.378	0	1
c 1945 1979	382	.18	.316	0	1
c 1980 1999	382	.189	.303	0	1.01
c post2000	382	.159	.291	0	1
c unknown	382	.091	.205	0	.82
p mgezw	382	.845	.262	0	1
p 1gezw	382	.155	.262	0	1
p huurwon	382	.673	.195	.02	1
p koopwon	382	.321	.194	0	.97
p bewndw	382	.924	.064	.61	1
p leegsw	382	.076	.064	0	.39
p opp land	382	.896	.129	.077	1
p opp water	382	.106	.128	0	.885
af artspr	382	526.371	440.857	100	5700
af ziek i	382	2264.23	1152.643	200	9700
af ziek e	382	2765.013	1312.516	200	10200
af superm	382	556.919	453.2	100	5700
af warenh	382	1650.653	1074.523	200	8500
af cafe	382	587.99	703.329	0	6900
af restau	382	336.292	333.102	0	2500
af hotel	382	835.509	725.701	0	3600
af kdv	382	369.974	267.256	100	3300
af ondbas	382	559.53	329.058	100	3200
af ondvrt	382	985.901	665.657	100	6900
af brandw	382	1761.097	946.819	100	6200
af oprith	382	2188.512	1054.024	300	5600
af treinst	382	2528.721	1621.343	300	10200
af overst	382	3671.54	1837.708	300	11600
af zwemb	382	2032.898	1198.5	200	9400
af ijsbaan	382	6806.789	2683.068	500	13200
af biblio	382	1266.057	705.028	200	7300
af zonbnk	382	1561.358	1545.885	100	7900
af attrac	382	3191.906	1775.064	300	11200
af museum	382	2428.721	1894.511	200	9000
bev dichth	382	13569.79	7964.801	34	35903

6

oad	382	5850.543	3211.492	28	12389
sted	382	1.23	.626	1	5
p mgz hh	382	.453	.14	.02	.81
p eenp hh	382	.547	.14	.19	.98
p west al	382	.197	.079	.04	.53
p n w al	382	.308	.197	.03	.87
p laaginkh	382	.496	.164	0	.98
p hooginkh	382	.184	.121	0	.57
p socminh	382	.12	.062	0	.38
p nietact	382	.63	.098	0	.88
ao uit tot	382	88.616	73.477	0	390
wwb uittot	382	108.486	114.29	0	560
a soz ow	382	253.708	192.427	0	1100
g wodief	382	5.248	4.358	0	38
g vernoo	382	6.034	7.192	0	56
g gewsek	382	13.379	31.782	0	376
c bev dichth	382	2	0.819	1	3

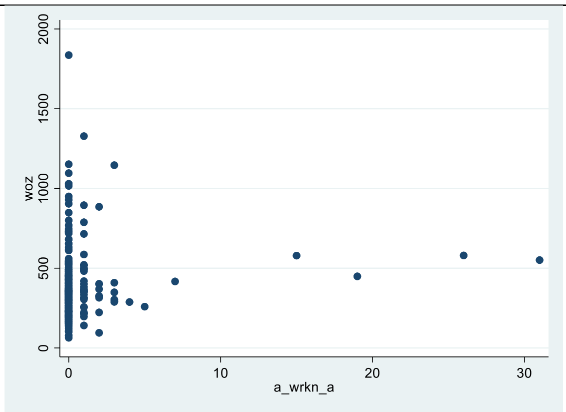
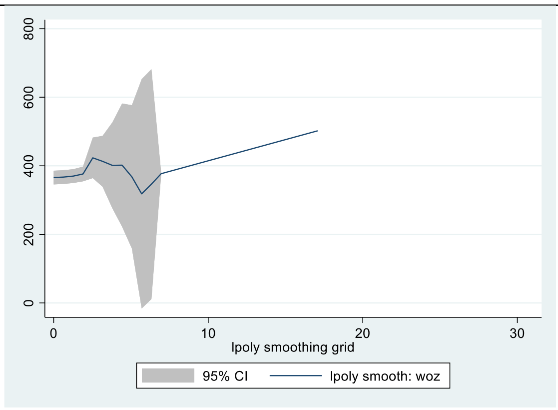
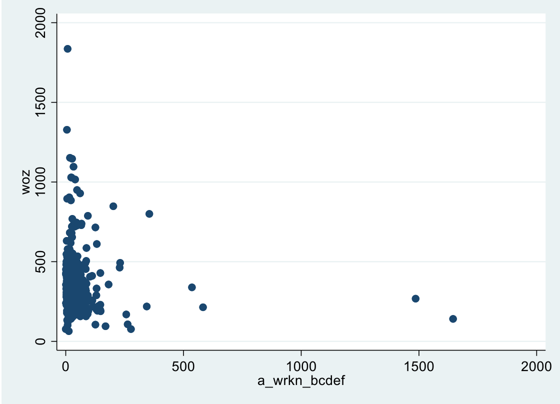
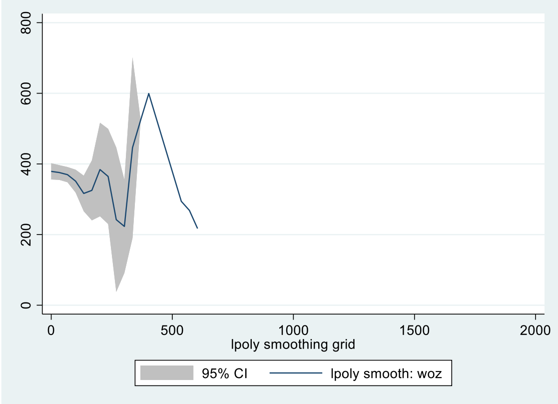
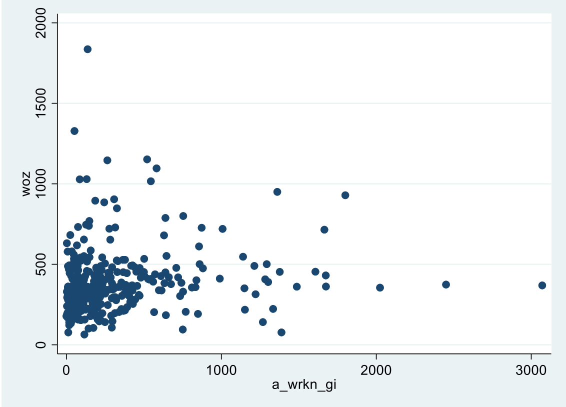
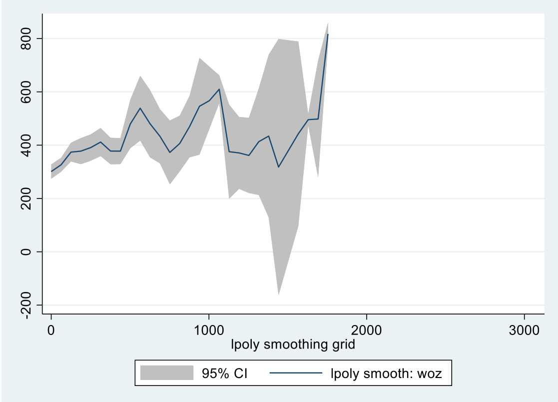
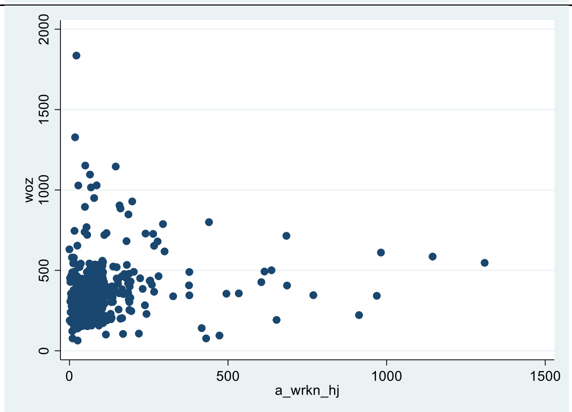
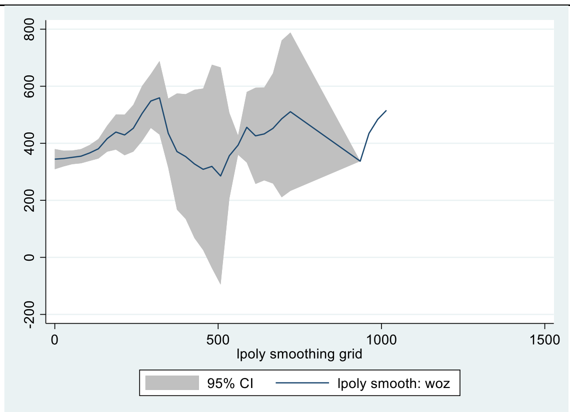
Appendix 8 Descriptive statistics of other variants of the employment sector variables (numbers and share)

Variable	Obs	Mean	Std. Dev.	Min	Max
Employees in merged employment sectors (number and share)					
a wrkn a	382	.454	2.483	0	31
a wrkn bcdef	382	52.423	125.038	0	1645
a wrkn gi	382	284.862	389.461	0	3071
a wrkn hj	382	110.99	158.158	0	1309
a wrkn kl	382	105.222	643.699	0	9678
a wrkn mn	382	269.164	400.695	2	4826
a wrkn opq	382	275.825	452.368	1	5366
a wrkn rstu	382	116.065	113.747	1	992
p wrkn a	382	.002	.014	0	.211
p wrkn bcdef	382	.053	.067	0	.737
p wrkn gi	382	.215	.152	0	.816
p wrkn hj	382	.099	.072	0	.661
p wrkn kl	382	.039	.082	0	.948
p wrkn mn	382	.219	.109	.006	.633
p wrkn opq	382	.255	.172	.003	.955
p wrkn rstu	382	.119	.07	.003	.539
Businesses in merged employment sectors (number and share)					
a vest a	382	.253	1.254	0	15
a vest bcdef	382	22.888	18.334	0	124
a vest gi	382	55.781	52.788	0	459
a vest hj	382	47.047	30.027	0	217
a vest kl	382	11.384	15.785	0	144
a vest mn	382	114.642	84.698	2	534
a vest opq	382	53.232	33.855	1	184
a vest rstu	382	64.93	48.885	1	280
p vest a	382	.002	.015	0	.212
p vest bcdef	382	.07	.051	0	.259
p vest gi	382	.144	.08	0	.613
p vest hj	382	.132	.045	0	.348
p vest kl	382	.028	.031	0	.235
p vest mn	382	.297	.083	.091	.515
p vest opq	382	.154	.061	.017	.442
p vest rstu	382	.173	.055	.023	.383
Employees in employment sectors as single entities (number and share)					
a wrkn a	382	.454	2.483	0	31
a wrkn b	382	.086	.679	0	11
a wrkn c	382	16.862	29.590	0	307
a wrkn d	382	7.06	87.369	0	1632
a wrkn e	382	3.125	56.156	0	1099
a wrkn f	382	25.29	42.083	0	579
a wrkn g	382	163.548	238.188	0	1871
a wrkn h	382	25.619	70.672	0	884
a wrkn i	382	121.313	201.273	0	1424
a wrkn j	382	85.371	137.755	0	1292
a wrkn k	382	85.982	616.519	0	9653
a wrkn l	382	19.24	56.762	0	897
a wrkn m	382	217.102	358.683	2	4664
a wrkn n	382	52.063	122.11	0	1556
a wrkn o	382	33.614	133.066	0	1211
a wrkn p	382	89.634	148.864	0	1788
a wrkn q	382	152.577	350.149	0	5312
a wrkn r	382	80.044	93.114	0	932

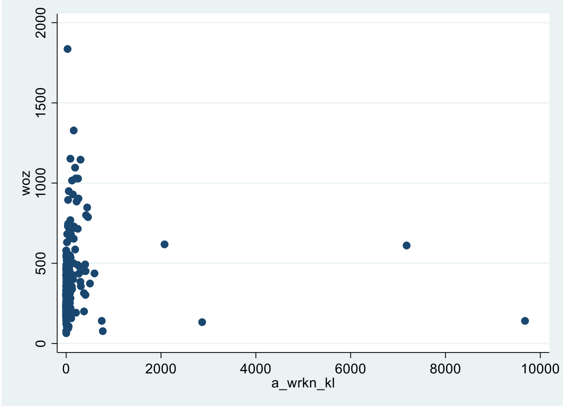
a wrkn s	382	35.632	43.865	0	513
a wrkn u	382	.389	2.951	0	36
p wrkn a	382	.002	.014	0	.211
p wrkn b	382	0	0	0	.008
p wrkn c	382	.017	.041	0	.714
p wrkn d	382	.001	.013	0	.186
p wrkn e	382	.002	.028	0	.545
p wrkn f	382	.033	.04	0	.301
p wrkn g	382	.127	.106	0	.641
p wrkn h	382	.031	.052	0	.572
p wrkn i	382	.089	.094	0	.635
p wrkn j	382	.068	.059	0	.654
p wrkn k	382	.026	.076	0	.947
p wrkn l	382	.013	.024	0	.313
p wrkn m	382	.172	.101	.003	.595
p wrkn n	382	.047	.057	0	.561
p wrkn o	382	.021	.073	0	.723
p wrkn p	382	.09	.086	0	.658
p wrkn q	382	.144	.132	0	.946
p wrkn r	382	.084	.066	0	.532
p wrkn s	382	.034	.023	0	.133
p wrkn u	382	0	.001	0	.018

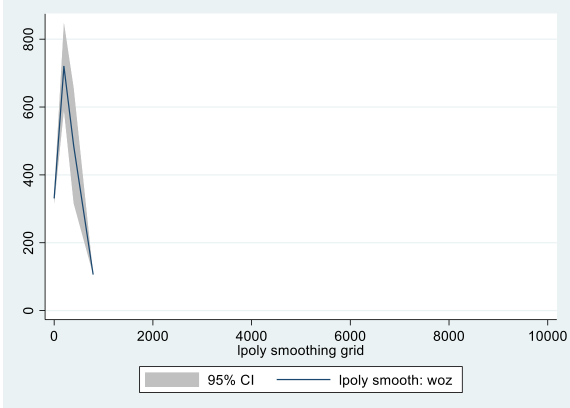
Each variable beginning with *a* indicates the number of employees or businesses, whereas *p_* indicates the percentages of employees or businesses.

Appendix 9 Correlation between the employment sectors and the average property price per neighbourhood

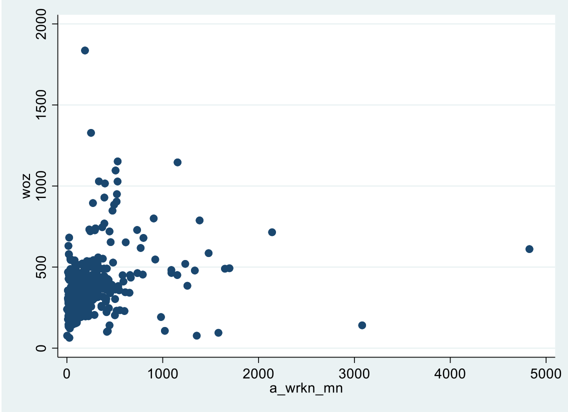
<table>
<tr><th>Sector</th><th>Scatter plot</th><th>Local polynomial line with smoothened line and 95% confidence interval</th></tr>
<tr><td>A</td><td>
</td><td>
</td></tr>
<tr><td>BCDEF</td><td>
</td><td>
</td></tr>
<tr><td>GI</td><td>
</td><td>
</td></tr>
<tr><td>HJ</td><td>
</td><td>
</td></tr>
</table>

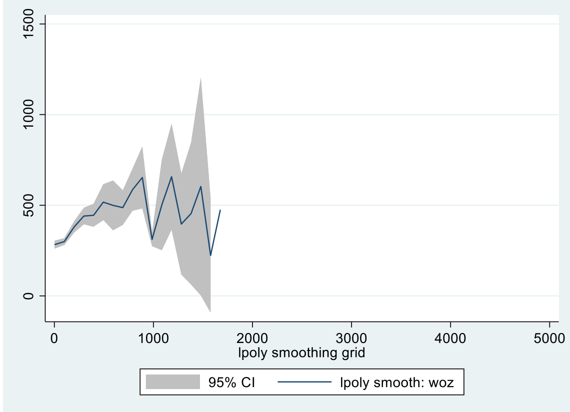
KL



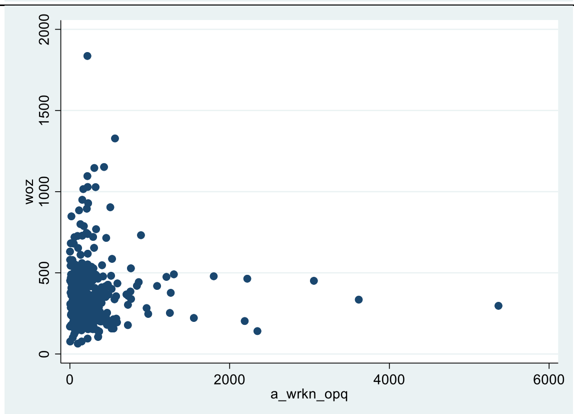


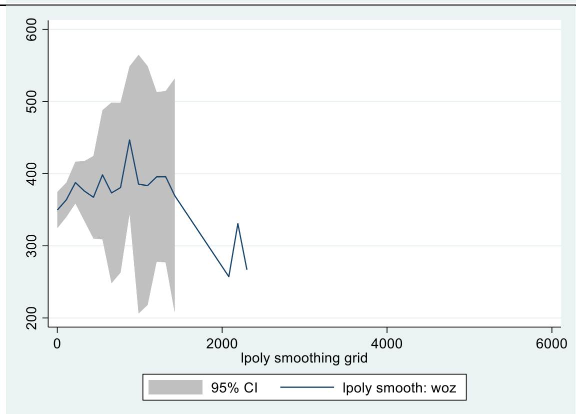
MN



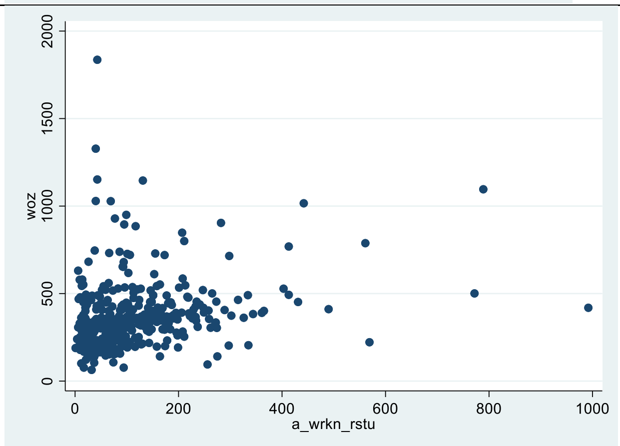


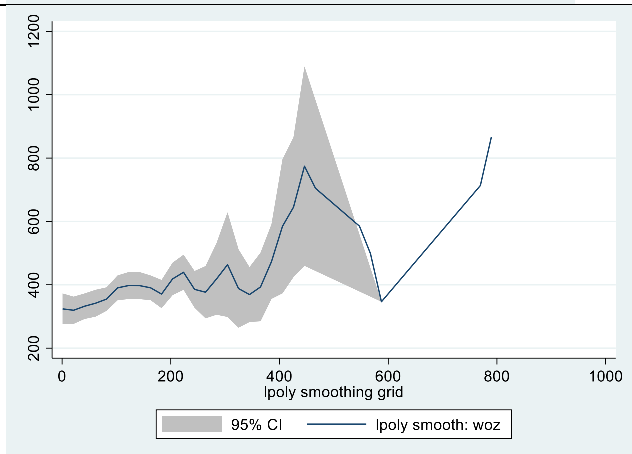
OPQ





RSTU





Appendix 10 Regression models (complete table)

VARIABLES	(1) Diversity index (DI)	(2a) + Average DI	(2b) + Weighted average DI	(3) + Employment sectors	(4) + Housing charact.	(5) + Spatial charact.	(6) + Social environment
Diversity index	0.227***	0.115***	0.164***	0.162***	0.0350*	0.0271	0.00242
	(0.0427)	(0.0361)	(0.0412)	(0.0435)	(0.0203)	(0.0189)	(0.0136)
Average diversity index		0.317***		0.298***	0.164***	0.104***	0.0740***
		(0.0448)		(0.0472)	(0.0351)	(0.0241)	(0.0201)
Weighted average diversity index			0.185***				
			(0.0451)				
Employees sector A (x100)				0.862*	-0.0610	1.095	0.596
				(0.479)	(0.294)	(0.737)	(0.560)
Employees sector BCDEF (x100)				-0.0711***	-0.0221	-0.0131	-0.00741
				(0.0235)	(0.0164)	(0.0123)	(0.00627)
Employees sector GI (x100)				-0.000783	-0.00187	-0.00256	-0.00221
				(0.00489)	(0.00301)	(0.00188)	(0.00205)
Employees sector HJ (x100)				-0.0361*	-0.00278	-0.00568	0.00386
				(0.0206)	(0.00801)	(0.00703)	(0.00507)
Employees sector KL (x100)				0.00700	0.00135	0.00118	0.000950
				(0.00536)	(0.00370)	(0.00265)	(0.00127)
Employees sector MN (x100)				-0.00610	0.00134	-0.000351	-0.00324
				(0.00980)	(0.00490)	(0.00386)	(0.00255)
Employees sector OPQ (x100)				-0.00357	0.00392*	0.00186	0.00346
				(0.00382)	(0.00214)	(0.00239)	(0.00251)
Employees sector RSTU (x100)				0.0184	0.00569	0.00588	0.0138*
				(0.0219)	(0.01000)	(0.00690)	(0.00699)
Homes					-1.59e-05	-1.17e-05	-0.000131***
					(2.33e-05)	(2.32e-05)	(3.27e-05)
avg_opp					0.000298	0.000179	0.000521
					(0.000755)	(0.000720)	(0.000552)
c_51_75					0.456***	0.576***	0.294***
					(0.127)	(0.0929)	(0.0774)
c_76_100					0.727***	0.931***	0.490***
					(0.126)	(0.0916)	(0.0919)
c_101_150					0.983***	1.019***	0.521***
					(0.185)	(0.127)	(0.101)
c_151_250					1.609***	1.783***	1.200***
					(0.273)	(0.224)	(0.234)
c_251_grot					2.522***	2.745***	2.586***
					(0.518)	(0.319)	(0.320)
c_pre1900					0.532***	0.300***	0.0764
					(0.0762)	(0.0712)	(0.0715)
c_1900_1944					0.464***	0.267***	0.0597
					(0.0650)	(0.0651)	(0.0490)
c_1945_1979					-0.0447	-0.116*	-0.0752*
					(0.0607)	(0.0597)	(0.0445)
c_1980_1999					-0.00703	-0.0328	-0.0414
					(0.0674)	(0.0609)	(0.0458)
c_unknown					0.367***	0.0504	-0.0157
					(0.107)	(0.0908)	(0.0627)
p_mgezw					0.172**	-0.139**	-0.0587

	(0.0707)	(0.0588)	(0.0455)
p_koopwon	0.386***	0.228**	-0.168*
	(0.125)	(0.0903)	(0.0896)
p_leegsw	0.358	0.289	0.688***
	(0.322)	(0.193)	(0.199)
p_opp_water		-0.0142	-0.100*
		(0.0980)	(0.0571)
af_artspr		9.08e-05	4.45e-05
		(5.70e-05)	(4.12e-05)
af_ziek_i		-5.22e-05*	-1.98e-05
		(3.14e-05)	(1.81e-05)
af_ziek_e		2.45e-05*	-7.68e-06
		(1.45e-05)	(1.17e-05)
af_superm		-1.04e-05	-5.01e-06
		(5.59e-05)	(3.62e-05)
af_warenh		1.02e-05	1.91e-05
		(2.72e-05)	(1.34e-05)
af_cafe		7.24e-05	7.75e-06
		(4.56e-05)	(2.26e-05)
af_restau		-2.98e-05	2.74e-06
		(7.49e-05)	(4.67e-05)
af_hotel		5.58e-05*	2.71e-05
		(3.16e-05)	(2.04e-05)
af_kdv		-0.000148*	-8.52e-05
		(8.60e-05)	(5.50e-05)
af_ondbas		8.52e-05**	6.24e-05*
		(4.25e-05)	(3.28e-05)
af_ondvrt		-4.16e-05	-2.53e-05
		(2.65e-05)	(1.66e-05)
af_brandw		-3.16e-05	1.33e-06
		(1.92e-05)	(1.17e-05)
af_oprith		-8.62e-06	1.81e-06
		(2.16e-05)	(1.27e-05)
af_treinst		3.16e-05**	1.75e-05*
		(1.45e-05)	(9.66e-06)
af_overst		-6.10e-05***	-4.64e-05***
		(1.14e-05)	(7.97e-06)
af_zwemb		3.18e-05	2.28e-05**
		(2.04e-05)	(1.14e-05)
af_ijsbaan		-8.81e-06	-1.59e-05***
		(6.37e-06)	(4.03e-06)
af_biblio		-2.20e-05	3.88e-07
		(3.28e-05)	(1.90e-05)
af_zonbnk		-4.72e-05***	-2.69e-05***
		(1.55e-05)	(8.28e-06)
af_attrac		-4.15e-05**	4.07e-06
		(1.63e-05)	(1.06e-05)
af_museum		-4.29e-05**	
		(1.90e-05)	
bev_dichth			5.47e-06**
			(2.20e-06)
oad			2.66e-05***
			(5.03e-06)
sted			-0.0450
			(0.0315)

p_west_al							-0.0882 (0.217)
p_n_w_al							-0.474*** (0.100)
p_laaginkh							-0.844*** (0.173)
p_hooginkh							0.581*** (0.200)
p_socminh							0.715** (0.278)
p_nietact							0.0658 (0.191)
ao_uit_tot							0.000626*** (0.000211)
a_soz_ow							0.000199** (7.64e-05)
g_wodief							-0.00338 (0.00208)
g_vernoo							5.78e-05 (0.00167)
g_gewsek							-0.000405 (0.000467)
Constant	12.10*** (0.111)	11.54*** (0.133)	11.73*** (0.144)	11.54*** (0.141)	11.09*** (0.174)	12.13*** (0.170)	12.58*** (0.252)
Observations	382	382	382	382	382	382	382
R-squared	0.248	0.395	0.317	0.440	0.824	0.898	0.951

Note: The dependent variable is the log transformed average property price per neighbourhood. The standard errors are clustered on a greater neighbourhood scale (wk_code), adding up to 96 neighbourhoods. Robust standard errors in parentheses. All models include a constant and error term. *** p<0.01, ** p<0.05, * p<0.10. The dummy variables are: “c_51_75 – c_251_grot”, “c_pre_1900 – c_unknown”, “p_mgezw”, “p_koopwon”, “p_leegsw”, “p_opp_water”, “p_west_al – p_n_w_al”, “p_laaginkh - p_hooginkh”, and “p_nietact”, see Appendix 4 for a further elaboration of the dummy variables.

Appendix 11 Elaboration of the remaining significant variables in Appendix 10

This appendix complements Section 4.5 and elaborates further on significant coefficients of the control variables as presented in Appendix 10.

Housing characteristics

The share of owner-occupied properties (*p*_koopwon) also exhibits significant results. This variable is a dummy variable and is compared with rental housing. Within models 4 and 5 in Appendix 10, a higher share of owner-occupied homes comes with higher average property prices in neighbourhoods compared to rental homes, being significant at 99% and 95%, respectively. Model 6 had negative sign, whereas an increase in owner-occupied homes is associated with lower property prices. The significance level is lower compared to models 4 and 5, indicating a less strong association with the average property prices in the neighbourhood, but there is no apparent cause for this.

The variable that indicates the percentage of vacant homes (p_leegsw) exhibits a positive sign and is significant at the 99% confidence interval in regression model 6. Based on the study of Lafferty and Frech (1978), vacancy was expected to be related to lower property prices. It is somewhat surprising that this variable is positive. It would be more likely that a higher share of vacancy would reflect less attractive neighbourhoods, thus being associated with lower property prices. However, due to the tense housing market, this might not be the case. Another possible explanation is that housing at the lower end of the market sells more easily than high-end housing. Therefore, vacant homes might reflect neighbourhoods with more housing in a higher segment and therefore relate to higher average property prices in the neighbourhood.

Spatial characteristics

As mentioned in earlier literature, proximity to water increases property prices (Daams et al., 2016; Rouwendal, Van Marwijk, and Levkovich, 2014). The average water surface (*p_opp_water*) in each neighbourhood is 10.6%, as presented in Appendix 7. This variable is a dummy variable and the reference variable is the share of land area (*p_opp_land*). An increase in the share of water surface was expected to be associated with higher average property prices. However, the results reveal otherwise, whereas model 6 is significant at 90%, indicating that an increase in the share of water has a negative association with the average property prices in the neighbourhoods. The city centre has many canals with considerably high housing prices, it was expected that water is associated with higher housing prices, but this does not appear to be true. Amsterdam also has many other water bodies, and thus appears to be associated with lower property prices for Amsterdam as a city.

Interesting results were found for distance to education. It appears that having elementary schools (*af_ondbas*) close by is not positively associated with the average property prices: the further the average distance to primary schools, the greater the average property prices in the neighbourhood. In models 5 and 6 in Appendix 10, this is significantly different from zero at the 95% and 90% confidence interval. Secondary schools (*af_ondvrt*), on the other hand, are associated with higher property prices if these schools are in proximity, and this variable is significant at the 99% confidence interval. Existing literature suggested a negative association from schools in direct vicinity (Koster and Rouwendal, 2012). Since there are more elementary schools in general and the association is measured per neighbourhood, this might indicate why elementary schools are negatively associated and secondary schools positively.

Furthermore, the distance to a major transfer station is appreciated by residents. An increase in distance to such a station (*af_overst*) has a negative association with the average property price in the neighbourhood. For both models, this variable is significant at the 99% confidence level. Thus, it can be stated that Amsterdam residents would like to have this facility in proximity. By contrast, an increasing distance to train stations (*af_treinst*) is positively associated at the 95 and 90% confidence interval in model 5 and model 6, respectively. Amsterdam has a large subway system throughout the city, but this is not taken into account since this variable only measures regular train stations. There are 11 train stations in Amsterdam, and except for Amsterdam Central Station, no other stations are located in or near the city centre. Neighbourhoods in the city centre are therefore located further from train stations, whereas neighbourhoods at the peripheries of Amsterdam are located closer to train stations. Considering Appendix 2a, property prices in the city centre towards Amsterdam South are generally higher compared with those at the peripheries. This might explain why a greater distance to a train station is associated with higher average property prices.

In addition, average house prices in neighbourhoods and distance to swimming pools (*af_zwem*) exhibits a significant relationship at the 95% confidence interval. Having a swimming pool further away is associated with higher property prices. On the other hand, the presence of an ice rink (*af_ijsbaan*) is negatively associated with property prices as the distance increases. Since Amsterdam only has one official ice rink, this variable is not representative despite its significance. A variable that exhibits a significant association at the 99% confidence interval is the distance to a solarium (*solarium*). As the distance to a solarium increases, there is a negative association with the average property price in the neighbourhood. Thus, a solarium in proximity is positively associated with property prices. A solarium is a specific type of amenity, it is likely that these solariums locate in in areas with higher property prices, whereas it is less likely that the reason for higher property prices comes from the presence of these solariums.

The final variable among the spatial characteristics is distance to museums (*af_museum*). As mentioned in Appendix 4, this variable is not applied in regression model 6 since it exhibited multicollinearity due to exceeding the critical value of the variance inflation factor (VIF). Moreover, this variable had the highest VIF of all variables in model 5, almost reaching the upper limit of 10 with a value of 9.55. Since this value was still within the boundaries of no multicollinearity, the coefficient was assumed to reflect the true value. As the distance to museums increased, average property prices within the neighbourhood are significantly associated with lower average property prices, indicating that residents would like to have a museum nearby. Noteworthy is that most museums in Amsterdam are concentrated in or nearby the higher priced neighbourhoods. This might explain why proximity to museums is associated with higher property prices.

<u>Social environment</u>

The final regression model adds a vector of variables relating to the social environment. The first variable is population density within the neighbourhood, measured in inhabitants per square kilometre. The coefficient is significantly different from zero at the 95% confidence interval, indicating that a greater population density is associated with higher average property values within the neighbourhood. A similar variable is the average address density (*OAD*).[7] The coefficient of this variable is also significantly different at the 99% confidence interval, whereas a greater average address density is also positively associated with the average property prices within the neighbourhood. Hypothesis 4 focused on the association between property values and a greater diversity of employment sectors within the neighbourhood. The average address density does not focus on the diversity of employment sectors, but only focusses on human activities, whereas an increase in activity leads to higher average property prices in the neighbourhood and is not contradictory with hypothesis 4.

The share of non-Western immigrants (*p_n_w_al*) and the average house prices in the neighbourhood are strongly associated. A higher share of non-Western immigrants is associated with higher average property values and is significant at the 99% confidence interval. This variable is also based on percentages and to measure the association of a 1%

[7] The OAD aims to indicate the degree of concentration of human activities, for example the combination of living, working, going to school, shopping, and going out (CBS, 2020). Each individual address has an OAD-value, indicating the average number of addresses per square kilometres, whereas all activity is measured within a radius of one kilometre. For each neighbourhood, an average OAD-value is calculated and is based on the average of all individual addresses.

increase, the coefficient is divided by 100. A 1% greater share of non-Western immigrants is negatively associated with 0.47% lower average property prices in the neighbourhood.

Furthermore, the income of households is also associated with the average property prices in the neighbourhood. The share of low-income households (*p_laaginkh*) and high-income households (*p_hooginkh*) are both significantly different from zero at the 99% confidence interval. These variables are dummy variables as well, and the reference category is the middle-income households. For a 1% greater share of low- and high-income households, the coefficients are again divided by 100. A 1% higher share of low-income households is associated with 0.84% lower average property prices in the neighbourhood, compared with middle-income households. Moreover, a 1% greater share of high-income households is associated with 0.58% higher property prices in comparison with middle-income households. This makes sense since higher income households are more likely to buy more expensive houses compared to middle- or low-income households.

The variable that indicates the share of households at or below the social minimum is also associated with average property prices in Amsterdam neighbourhoods. This variable is significant at the 95% confidence interval. A 1% greater share of households at or below the social minimum is associated with 0.72% lower average property prices in the neighbourhood. Since these households are related to low-income households, a more obvious result would be to have lower average property prices.

Also, households that receive government benefits are associated with lower average property prices. A greater number of households in the neighbourhood receiving benefits via the General Disability Act (*ao_uit_tot*) and the General Old Age Pension Act (*a_soz_ow*) exhibit positive associations with the average property prices. What both types of residents have in common is that most are likely to be unemployed, what in general would be associated with a lower income. There is no obvious explanation for this positive association, since an opposite association would have made more sense.

Appendix 12 Nonlinear figures

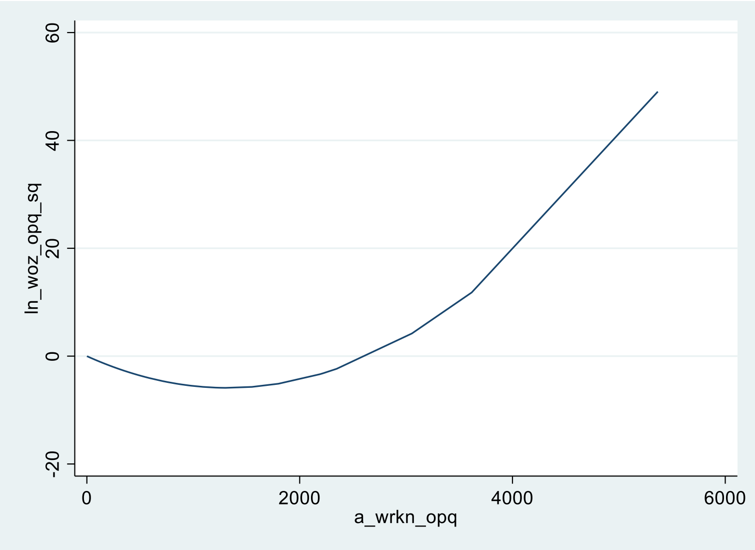


Figure 11.1: The non-linear association between average property prices within a neighbourhood and more employees in sector OPQ (government, education and healthcare).

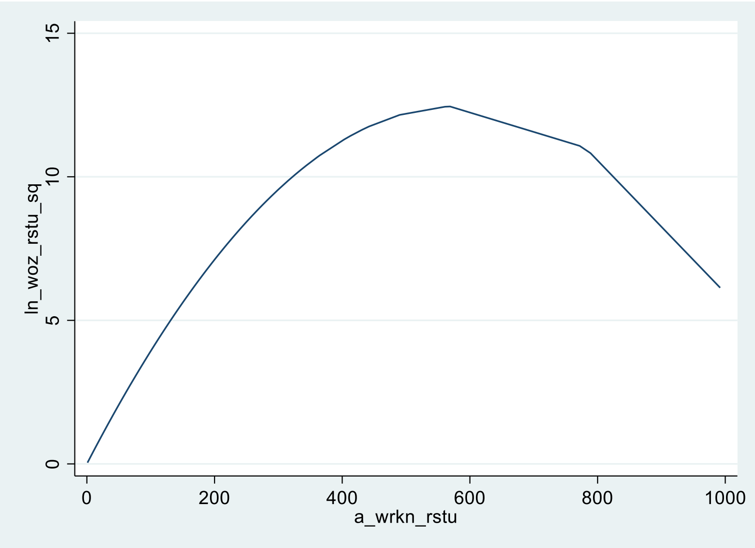


Figure 11.2: The non-linear association between average property prices within a neighbourhood and more employees in sector RSTU (culture, recreation and other services).

Appendix 13 Calculation Chow test

The formula for calculating the Chow test according to Brooks and Tsolacos (2010) is:

$$F = \frac{RSS - (RSS_1 + RSS_2 + RSS_3)}{RSS_1 + RSS_2 + RSS_3} * \frac{T - 3k}{3k - k}$$

F	:	F-statistic
RSS	:	Residual sum of squares for the whole sample
RSS_1	:	Residual sum of squares of the low-densely populated neighbourhood
RSS_2	:	Residual sum of squares of the middle-densely populated neighbourhood
RSS_3	:	Residual sum of squares of the high-densely populated neighbourhood
T	:	Number of observations
k	:	Number of regressors

The outcome of the Chow test is:

$$F = \frac{4.064 - (1.566 \ + 0.556 + 0.182)}{1.566 + 0.556 + 0.182} * \frac{382 - (3 * 60)}{(3 * 60) - 60} = 1.286$$

Critical value F(120,202) = 1.221 at the significance level of 5%.

The Chow F-test is slightly larger than the critical value, thereby rejecting the null hypothesis that the intercepts and slopes of low-, middle- and high-densely populated neighbourhoods are identical.

Appendix 14 Regression models Chow test

VARIABLES	**Pooled** Regression model 6	**Group 1** Lowest population density group	**Group 2** Middle population density group	**Group 3** Highest population density group
di	0.00242	-0.00498	-0.00713	0.00369
	(0.0123)	(0.0240)	(0.0184)	(0.0265)
avg_di	0.0740***	0.0780*	0.0568*	0.0284
	(0.0194)	(0.0454)	(0.0331)	(0.0225)
a_wrkn_a_x100	0.596	-0.145	5.193**	0.751
	(0.527)	(0.821)	(2.525)	(0.678)
a_wrkn_bcdef_x100	-0.00741	0.00531	0.00605	-0.0297
	(0.00611)	(0.0129)	(0.0187)	(0.0375)
a_wrkn_gi_x100	-0.00221	-0.00409	-0.00425	-0.00696
	(0.00219)	(0.00697)	(0.00359)	(0.00477)
a_wrkn_hj_x100	0.00386	0.00907	-0.00768	0.00723
	(0.00459)	(0.0127)	(0.00887)	(0.0113)
a_wrkn_kl_x100	0.000950	-0.00531	-0.00319	0.0257
	(0.00150)	(0.00356)	(0.00429)	(0.0447)
a_wrkn_mn_x100	-0.00324	-0.00407	0.00815	0.00592
	(0.00283)	(0.00572)	(0.00617)	(0.00683)
a_wrkn_opq_x100	0.00346	0.00457	-0.00114	-0.00175
	(0.00253)	(0.00355)	(0.00294)	(0.00219)
a_wrkn_rstu_x100	0.0138*	0.0163	0.0144	0.00635
	(0.00749)	(0.0158)	(0.0138)	(0.0176)
homes	-0.000131***	-0.000167*	-0.000134**	-4.39e-05
	(3.68e-05)	(8.52e-05)	(5.62e-05)	(3.92e-05)
avg_opp	0.000521	0.00102	0.00432	0.000854
	(0.000621)	(0.00109)	(0.00309)	(0.00330)
c_51_75	0.294***	0.571***	-0.235	0.0774
	(0.0744)	(0.143)	(0.147)	(0.0846)
c_76_100	0.490***	0.480***	0.0445	0.330*
	(0.0904)	(0.167)	(0.209)	(0.185)
c_101_150	0.521***	0.581***	-0.227	0.442
	(0.101)	(0.205)	(0.332)	(0.288)
c_151_250	1.200***	1.597***	0.289	0.696
	(0.229)	(0.494)	(0.553)	(0.802)
c_251_grot	2.586***	2.137***	0.653	1.669
	(0.339)	(0.600)	(1.655)	(2.181)
c_pre1900	0.0764	0.0433	0.204	0.0339
	(0.0684)	(0.196)	(0.152)	(0.0780)
c_1900_1944	0.0597	0.0185	0.0153	0.0736
	(0.0480)	(0.113)	(0.0865)	(0.0665)
c_1945_1979	-0.0752	-0.0467	-0.116	-0.219**
	(0.0508)	(0.0998)	(0.0724)	(0.110)
c_1980_1999	-0.0414	0.0348	-0.0552	0.00526
	(0.0439)	(0.104)	(0.0695)	(0.0682)
c_unknown	-0.0157	0.0125	-0.0559	-0.0485
	(0.0623)	(0.216)	(0.154)	(0.120)
p_mgezw	-0.0587	-0.0948	0.0748	0.253

	(0.0472)	(0.0925)	(0.0763)	(0.210)
p_koopwon	-0.168**	-0.213	-0.230*	-0.364***
	(0.0792)	(0.142)	(0.130)	(0.132)
p_leegsw	0.688***	0.607	0.208	0.565*
	(0.217)	(0.429)	(0.334)	(0.292)
p_opp_water	-0.100*	-0.0761	-0.122	-0.00271
	(0.0546)	(0.196)	(0.131)	(0.0769)
af_artspr	4.45e-05	3.25e-05	9.80e-05	6.26e-05
	(4.55e-05)	(6.31e-05)	(6.32e-05)	(6.13e-05)
af_ziek_i	-1.98e-05	-1.61e-05	-2.24e-05	-1.01e-05
	(1.41e-05)	(7.86e-05)	(2.48e-05)	(2.84e-05)
af_ziek_e	-7.68e-06	-1.15e-06	2.50e-07	1.31e-05
	(1.01e-05)	(7.31e-05)	(1.90e-05)	(2.58e-05)
af_superm	-5.01e-06	1.46e-05	-7.85e-05	-1.10e-05
	(4.20e-05)	(6.13e-05)	(6.57e-05)	(5.57e-05)
af_warenh	1.91e-05	-9.27e-06	-4.15e-06	-3.32e-06
	(1.20e-05)	(3.15e-05)	(2.27e-05)	(1.74e-05)
af_cafe	7.75e-06	4.69e-05	-3.09e-05	-5.37e-06
	(1.96e-05)	(4.93e-05)	(3.58e-05)	(6.91e-05)
af_restau	2.74e-06	2.53e-05	-0.000188*	-0.000174
	(4.16e-05)	(6.52e-05)	(0.000103)	(0.000114)
af_hotel	2.71e-05	5.98e-06	8.58e-05***	-8.01e-05*
	(1.83e-05)	(4.16e-05)	(3.18e-05)	(4.24e-05)
af_kdv	-8.52e-05	-0.000200**	-9.43e-06	-3.68e-05
	(5.56e-05)	(9.81e-05)	(0.000114)	(8.72e-05)
af_ondbas	6.24e-05*	0.000164**	-3.07e-05	-1.03e-05
	(3.30e-05)	(7.74e-05)	(7.02e-05)	(4.88e-05)
af_ondvrt	-2.53e-05	-2.09e-05	5.75e-06	-4.23e-05*
	(1.61e-05)	(3.80e-05)	(3.44e-05)	(2.43e-05)
af_brandw	1.33e-06	5.14e-05	6.38e-06	2.06e-05
	(1.07e-05)	(3.18e-05)	(2.04e-05)	(1.76e-05)
af_oprith	1.81e-06	-2.17e-05	9.89e-06	1.27e-06
	(1.08e-05)	(2.73e-05)	(1.82e-05)	(1.76e-05)
af_treinst	1.75e-05**	4.24e-05*	2.51e-05	1.92e-05
	(8.28e-06)	(2.22e-05)	(1.59e-05)	(2.05e-05)
af_overst	-4.64e-05***	-3.00e-05	-4.03e-05**	-1.54e-05
	(7.81e-06)	(2.63e-05)	(1.63e-05)	(1.58e-05)
af_zwemb	2.28e-05**	4.86e-05**	5.75e-06	-1.25e-05
	(1.00e-05)	(2.32e-05)	(2.13e-05)	(1.75e-05)
af_ijsbaan	-1.59e-05***	-2.37e-05**	-1.13e-05*	-1.18e-05
	(3.54e-06)	(1.03e-05)	(6.33e-06)	(8.57e-06)
af_biblio	3.88e-07	-1.39e-05	1.62e-05	-4.21e-06
	(1.69e-05)	(4.58e-05)	(2.62e-05)	(2.50e-05)
af_zonbnk	-2.69e-05***	-3.90e-05**	-1.81e-05	7.35e-06
	(7.76e-06)	(1.59e-05)	(1.71e-05)	(2.08e-05)
af_attrac	4.07e-06	-2.54e-05	-1.50e-05	-2.12e-05
	(9.23e-06)	(2.46e-05)	(1.80e-05)	(1.96e-05)
bev_dichth	5.47e-06***	5.22e-06	-2.73e-06	1.31e-06
	(2.09e-06)	(1.24e-05)	(5.63e-06)	(2.17e-06)
oad	2.66e-05***	3.25e-05*	3.18e-05***	4.61e-06

	(4.36e-06)	(1.94e-05)	(1.09e-05)	(6.02e-06)
sted	-0.0450*	-0.0475	0.00231	-0.164
	(0.0250)	(0.0385)	(0.0470)	(0.325)
p_west_al	-0.0882	0.0110	-0.230	0.495
	(0.235)	(0.404)	(0.460)	(0.393)
p_n_w_al	-0.474***	-0.373	-0.537***	-0.512***
	(0.0931)	(0.232)	(0.180)	(0.155)
p_laaginkh	-0.844***	-0.831***	-0.965***	-0.468
	(0.169)	(0.199)	(0.352)	(0.338)
p_hooginkh	0.581***	0.877***	0.626	1.025**
	(0.174)	(0.312)	(0.435)	(0.481)
p_socminh	0.715**	1.113**	0.459	-0.707
	(0.300)	(0.491)	(0.632)	(0.480)
p_nietact	0.0658	0.00329	-0.112	-1.265***
	(0.184)	(0.242)	(0.394)	(0.265)
ao_uit_tot	0.000626***	0.000566	0.000635	0.000445
	(0.000226)	(0.000562)	(0.000414)	(0.000303)
a_soz_ow	0.000199***	8.71e-05	0.000169	4.89e-05
	(7.65e-05)	(0.000185)	(0.000138)	(0.000112)
g_wodief	-0.00338*	-0.00241	-0.00304	-0.00484
	(0.00203)	(0.00367)	(0.00372)	(0.00297)
g_vernoo	5.78e-05	0.000925	-0.00657*	0.00305
	(0.00178)	(0.00398)	(0.00365)	(0.00271)
g_gewsek	-0.000405	-7.76e-05	0.000727	0.000113
	(0.000378)	(0.000817)	(0.000596)	(0.000921)
Constant	12.58***	12.33***	13.11***	13.62***
	(0.250)	(0.325)	(0.449)	(0.466)
Observations	382	127	127	128
R-squared	0.951	0.962	0.983	0.974

Note: The dependent variable is the log transformed average property price per neighbourhood. The standard errors are clustered on a greater neighbourhood scale (wk_code), adding up to 96 neighbourhoods. Robust standard errors in parentheses. All models include a constant and error term. *** p<0.01, ** p<0.05, * p<0.10. The dummy variables are: “c_51_75 – c_251_grot”, “c_pre_1900 – c_unknown”, “p_mgezw”, “p_koopwon”, “p_leegsw”, “p_opp_water”, “p_west_al – p_n_w_al”, “p_laaginkh - p_hooginkh”, and “p_nietact”, see Appendix 4 for a further elaboration of the dummy variables.

Appendix 15 Stata syntax

```
*** 1	Import data-file
clear all
import excel "C:\Users\user\Documents\Dataset_CBS_LISA_Div_Ind.xlxs

*** 2		Log transforming WOZ-value
generate ln_woz = ln(woz_x1000)
hist woz_x1000
hist ln_woz

*** 3		Descriptive Statistics
asdoc sum ln_woz di_8 avg_di_8 w_di_ba_8 a_wrkn_a a_wrkn_bcdef a_wrkn_gi
	a_wrkn_hj a_wrkn_kl a_wrkn_mn a_wrkn_opq a_wrkn_rstu woningen
	avg_opp c_0_50 c_51_75 c_76_100 c_101_150
	c_151_250 c_251_grot avg_bwj c_pre1900 c_1900_1944
	c_1945_1979 c_1980_1999 c_post2000 c_unknown p_mgezw
	p_1gezw p_huurwon p_koopwon p_bewndw p_leegsw
	p_opp_land p_opp_water af_artspr af_ziek_i af_ziek_e af_superm
	af_warenh af_cafe af_restau af_hotel af_kdv af_ondbas
	af_ondvrt af_brandw af_oprith af_treinst af_overst
	af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac af_museum
	bev_dichth oad sted p_mgz_hh p_eenp_hh p_west_al p_n_w_al
	p_laaginkh p_hooginkh p_socminh p_nietact ao_uit_tot wwb_uittot
	a_soz_ow g_wodief g_vernoo g_gewsek c_bev_dichth_3

*** 4	Regression models standard (incl. creating table)
generate a_wrkn_a_x100 = a_wrkn_a /100
generate a_wrkn_bcdef_x100 = a_wrkn_bcdef /100
generate a_wrkn_gi_x100 = a_wrkn_gi /100
generate a_wrkn_hj_x100 = a_wrkn_hj /100
generate a_wrkn_kl_x100 = a_wrkn_kl /100
generate a_wrkn_mn_x100 = a_wrkn_mn /100
generate a_wrkn_opq_x100 = a_wrkn_opq /100
generate a_wrkn_rstu_x100 = a_wrkn_rstu /100

* MODEL 1 - ln(P_n) = a + B1D_n + e_n
reg ln_woz di_8, vce(cluster wk_code)
outreg2 using regmodels_standard.doc, replace ctitle (model 1)

* MODEL 2a - ln(P_n) = a + B1Dn + B2Da_n + e_n
* MODEL 2b - ln(P_n) = a + B1Dn + B2Dwa_n + e_n
reg ln_woz di_8 avg_di_8, vce(cluster wk_code)
outreg2 using regmodels_standard.doc, append ctitle (model 2a)
reg ln_woz di_8 w_di_ba_8, vce(cluster wk_code)
outreg2 using regmodels_standard.doc, append ctitle (model 2b)

* Model 3 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100,
vce(cluster wk_code)
outreg2 using regmodels_standard.doc, append ctitle (model 3)

* Model 4 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + e_n
```

```
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw,
vce(cluster wk_code)
outreg2 using regmodels_standard.doc, append ctitle (model 4)

* Model 5 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900	c_1900_1944 c_1945_1979
	c_1980_1999	c_unknown	p_mgezw		p_koopwon
	p_leegsw	p_opp_water	af_artspr	af_ziek_i	af_ziek_e
	af_superm	af_warenh	af_cafe	af_restau	af_hotel	af_kdv
	af_ondbas	af_ondvrt		af_brandw	af_oprith	af_treinst
	af_overst	af_zwemb	af_ijsbaan	af_biblio	af_zonbnk	af_attrac
	af_museum, vce(cluster wk_code)
outreg2 using regmodels_standard.doc, append ctitle (model 5)

* Model 6 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek,	vce(cluster
wk_code)
outreg2 using regmodels_standard.doc, append ctitle (model 6)

** 5		Checking for multicollinearity (VIF & correlation matrixes)
* MODEL 1 - ln(P_n) = a + B1D_n + e_n
quietly reg ln_woz di_8, vce(cluster wk_code)
estat vif
cor ln_woz di_8

* MODEL 2a - ln(P_n) = a + B1Dn + B2Da_n + e_n
* MODEL 2b - ln(P_n) = a + B1Dn + B2Dwa_n + e_n
quietly reg ln_woz di_8 avg_di_8, vce(cluster wk_code)
estat vif
cor ln_woz di_8 avg_di_8
quietly reg ln_woz di_8 w_di_ba_8, vce(cluster wk_code)
estat vif
cor ln_woz di_8 w_di_ba_8

* Model 3 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100,
vce(cluster wk_code)
```

```
estat vif
cor ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100

* Model 4 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp	c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw,
vce(cluster wk_code)
estat vif
cor ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp	c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw

* Model 5 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp	c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900	c_1900_1944 c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon
	p_leegsw	p_opp_water	af_artspr	af_ziek_i	af_ziek_e
	af_superm	af_warenh	af_cafe	af_restau	af_hotel	af_kdv
	af_ondbas	af_ondvrt	af_brandw	af_oprith	af_treinst
	af_overst	af_zwemb	af_ijsbaan	af_biblio	af_zonbnk	af_attrac
	af_museum, vce(cluster wk_code)
estat vif
cor ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp	c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900	c_1900_1944 c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon
	p_leegsw	p_opp_water	af_artspr	af_ziek_i	af_ziek_e
	af_superm	af_warenh	af_cafe	af_restau	af_hotel	af_kdv
	af_ondbas	af_ondvrt	af_brandw	af_oprith	af_treinst
	af_overst	af_zwemb	af_ijsbaan	af_biblio	af_zonbnk	af_attrac
	af_museum

* Model 6 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp	c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek,	vce(cluster
wk_code)
estat vif
```

```
cor ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek

*** 6		Testing OLS-assumptions
** OLS assumption 2 - heteroscedasticity
* MODEL 1 - ln(P_n) = a + B1D_n + e_n
quietly reg ln_woz di_8
estat hettest
rvfplot, yline(0)
* MODEL 2a - ln(P_n) = a + B1Dn + B2Da_n + e_n
* MODEL 2b - ln(P_n) = a + B1Dn + B2Dwa_n + e_n
quietly reg ln_woz di_8 avg_di_8
estat hettest
rvfplot, yline(0)
quietly reg ln_woz di_8 w_di_ba_8
estat hettest
rvfplot, yline(0)
* Model 3 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
estat hettest
rvfplot, yline(0)

* Model 4 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot		c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw
estat hettest
rvfplot, yline(0)

* Model 5 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900	c_1900_1944 c_1945_1979
	c_1980_1999	c_unknown	p_mgezw		p_koopwon
	p_leegsw	p_opp_water	af_artspr	af_ziek_i	af_ziek_e
	af_superm	af_warenh	af_cafe	af_restau	af_hotel	af_kdv
	af_ondbas	af_ondvrt		af_brandw	af_oprith	af_treinst
	af_overst	af_zwemb	af_ijsbaan	af_biblio	af_zonbnk	af_attrac
	af_museum
estat hettest
rvfplot, yline(0)
```

```
* Model 6 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp			c_51_75		c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999		c_unknown	p_mgezw		p_koopwon		p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe		af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek
estat hettest
rvfplot, yline(0)

** OLS assumption 4 - Endogeneity
* MODEL 1 - ln(P_n) = a + B1D_n + e_n
quietly reg ln_woz di_8
quietly reg di_8
predict res_m1, resid
 reg ln_woz di_8 res_m1
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 1)

* MODEL 2a - ln(P_n) = a + B1Dn + B2Da_n + e_n
* MODEL 2b - ln(P_n) = a + B1Dn + B2Dwa_n + e_n
quietly reg ln_woz di_8 avg_di_8
quietly reg di_8 avg_di_8
predict res_m2a, resid
quietly reg ln_woz di_8 res_m2a avg_di_8
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 2a)

quietly reg ln_woz di_8 w_di_ba_8
quietly reg di_8 w_di_ba_8
predict res_m2b, resid
quietly reg ln_woz di_8 res_m2b w_di_ba_8
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 2b)

* Model 3 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
quietly reg di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
predict res_m3, resid
quietly reg ln_woz di_8 res_m3 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100
a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100
a_wrkn_rstu_x100
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 3)

* Model 4 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp			c_51_75		c_76_100	c_101_150
	c_151_250	c_251_grot		c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999		c_unknown	p_mgezw		p_koopwon		p_leegsw
quietly reg di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
```

```
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
    woningen avg_opp c_51_75 c_76_100 c_101_150
    c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979
    c_1980_1999 c_unknown p_mgezw p_koopwon p_leegsw
predict res_m4, resid
quietly reg ln_woz di_8 res_m4 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100
a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100
a_wrkn_rstu_x100 woningen avg_opp c_51_75 c_76_100
    c_101_150 c_151_250 c_251_grot c_pre1900 c_1900_1944
    c_1945_1979 c_1980_1999 c_unknown p_mgezw p_koopwon
    p_leegsw
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 4)

* Model 5 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
    woningen avg_opp c_51_75 c_76_100 c_101_150
    c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979
    c_1980_1999 c_unknown p_mgezw p_koopwon
    p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e
    af_superm af_warenh af_cafe af_restau af_hotel af_kdv
    af_ondbas af_ondvrt af_brandw af_oprith af_treinst
    af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac
    af_museum
quietly reg di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
    woningen avg_opp c_51_75 c_76_100 c_101_150
    c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979
    c_1980_1999 c_unknown p_mgezw p_koopwon
    p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e
    af_superm af_warenh af_cafe af_restau af_hotel af_kdv
    af_ondbas af_ondvrt af_brandw af_oprith af_treinst
    af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac
    af_museum
predict res_m5, resid
. quietly reg ln_woz di_8 res_m5 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100
a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100
a_wrkn_rstu_x100 woningen avg_opp c_51_75 c_76_100
    c_101_150 c_151_250 c_251_grot c_pre1900 c_1900_1944
    c_1945_1979 c_1980_1999 c_unknown p_mgezw p_koopwon
        p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e
    af_superm af_warenh af_cafe af_restau af_hotel af_kdv
    af_ondbas af_ondvrt af_brandw af_oprith af_treinst
    af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac
    af_museum
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 5)

* Model 6 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
    woningen avg_opp c_51_75 c_76_100 c_101_150
    c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979
    c_1980_1999 c_unknown p_mgezw p_koopwon p_leegsw
    p_opp_water af_artspr af_ziek_i af_ziek_e af_superm af_warenh
    af_cafe af_restau af_hotel af_kdv af_ondbas af_ondvrt
```

```
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek
quietly reg di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek
predict res_m6, resid
quietly reg di_8 res_m6 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp		c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999	c_unknown	p_mgezw	p_koopwon	p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek
outreg2 using regmodels_standard_endogeneity.doc, append ctitle (model 6)

** OLS assumption 5 - normality
* MODEL 1 - ln(P_n) = a + B1D_n + e_n
quietly reg ln_woz di_8
predict res_model1, resid
kdensity res_model1, normal
swilk res_model1

* MODEL 2a - ln(P_n) = a + B1Dn + B2Da_n + e_n
* MODEL 2b - ln(P_n) = a + B1Dn + B2Dwa_n + e_n
quietly reg ln_woz di_8 avg_di_8
predict res_model2a, resid
kdensity res_model2a, normal
swilk res_model2a
quietly reg ln_woz di_8 w_di_ba_8
predict res_model2b, resid
kdensity res_model2b, normal
swilk res_model2b

* Model 3 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
predict res_model3, resid
kdensity res_model3, normal
swilk res_model3

* Model 4 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + e_n
```

```
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp			c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot		c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999		c_unknown	p_mgezw	p_koopwon		p_leegsw
predict res_model4, resid
kdensity res_model4, normal
swilk res_model4

* Model 5 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp			c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900	c_1900_1944 c_1945_1979
	c_1980_1999		c_unknown	p_mgezw		p_koopwon
	p_leegsw		p_opp_water	af_artspr	af_ziek_i	af_ziek_e
	af_superm	af_warenh	af_cafe		af_restau	af_hotel	af_kdv
	af_ondbas	af_ondvrt		af_brandw	af_oprith	af_treinst
	af_overst	af_zwemb	af_ijsbaan	af_biblio	af_zonbnk	af_attrac
	af_museum
predict res_model5, resid
kdensity res_model5, normal
swilk res_model5

* Model 6 - ln(P_n) = a + B1Dn + B2Da_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
quietly reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	woningen	avg_opp			c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999		c_unknown	p_mgezw	p_koopwon		p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe		af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek
predict res_model6, resid
kdensity res_model6, normal
swilk res_model6

*** 7		Robustness test 1: Businesses instead of employees (incl. creating table)
* Model 3 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + e_n
reg ln_woz di_8 avg_di_8 a_vest_a a_vest_bcdef a_vest_gi	a_vest_hj	a_vest_kl
	a_vest_mn	a_vest_opq	a_vest_rstu, vce(cluster wk_code)
outreg2 using regmodels_number_businesses.doc, append ctitle (model 3)

* Model 4 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + e_n
reg ln_woz di_8 avg_di_8 a_vest_a a_vest_bcdef a_vest_gi	a_vest_hj	a_vest_kl
	a_vest_mn	a_vest_opq	a_vest_rstu	woningen	avg_opp
	c_51_75		c_76_100	c_101_150	c_151_250	c_251_grot
	c_pre1900 c_1900_1944	c_1945_1979 c_1980_1999		c_unknown
	p_mgezw		p_koopwon		p_leegsw, vce(cluster wk_code)
outreg2 using regmodels_number_businesses.doc, append ctitle (model 4)

* Model 5 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + e_n
```

```
reg ln_woz di_8 avg_di_8 a_vest_a a_vest_bcdef a_vest_gi a_vest_hj a_vest_kl
	a_vest_mn a_vest_opq a_vest_rstu woningen avg_opp
	c_51_75 c_76_100 c_101_150 c_151_250 c_251_grot c_pre1900
	c_1900_1944 c_1945_1979 c_1980_1999 c_unknown p_mgezw
	p_koopwon p_leegsw p_opp_water af_artspr af_ziek_i
	af_ziek_e af_superm af_warenh af_cafe af_restau af_hotel
	af_kdv af_ondbas af_ondvrt af_brandw af_oprith af_treinst
	af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac
	af_museum, vce(cluster wk_code)
outreg2 using regmodels_number_businesses.doc, append ctitle (model 5)

* Model 6 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
reg ln_woz di_8 avg_di_8 a_vest_a a_vest_bcdef a_vest_gi a_vest_hj a_vest_kl
	a_vest_mn a_vest_opq a_vest_rstu woningen avg_opp
	c_51_75 c_76_100 c_101_150 c_151_250 c_251_grot c_pre1900
c_1900_1944 c_1945_1979 c_1980_1999 c_unknown p_mgezw p_koopwon
		p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e
	af_superm af_warenh af_cafe af_restau af_hotel af_kdv
	af_ondbas af_ondvrt af_brandw af_oprith af_treinst
	af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac
	bev_dichth oad sted p_west_al p_n_w_al p_laaginkh
	p_hooginkh p_socminh p_nietact ao_uit_tot a_soz_ow g_wodief
	g_vernoo g_gewsek, vce(cluster wk_code)
outreg2 using regmodels_number_businesses.doc, append ctitle (model 6)

*** 8 Robustness test 2: using 21 employment sectors instead of 8 (incl. creating table)
* Generating variables *100
generate a_wrkn_a_21_x100 = a_wrkn_a_21 /100
generate a_wrkn_b_x100 = a_wrkn_b /100
generate a_wrkn_c_x100 = a_wrkn_c /100
generate a_wrkn_d_x100 = a_wrkn_d /100
generate a_wrkn_e_x100 = a_wrkn_e /100
generate a_wrkn_f_x100 = a_wrkn_f /100
generate a_wrkn_g_x100 = a_wrkn_g /100
generate a_wrkn_h_x100 = a_wrkn_h /100
generate a_wrkn_i_x100 = a_wrkn_i /100
generate a_wrkn_j_x100 = a_wrkn_j /100
generate a_wrkn_k_x100 = a_wrkn_k /100
generate a_wrkn_l_x100 = a_wrkn_l /100
generate a_wrkn_m_x100 = a_wrkn_m /100
generate a_wrkn_n_x100 = a_wrkn_n /100
generate a_wrkn_o_x100 = a_wrkn_o /100
generate a_wrkn_p_x100 = a_wrkn_p /100
generate a_wrkn_q_x100 = a_wrkn_q /100
generate a_wrkn_r_x100 = a_wrkn_r /100
generate a_wrkn_s_x100 = a_wrkn_s /100
generate a_wrkn_t_x100 = a_wrkn_t /100
generate a_wrkn_u_x100 = a_wrkn_u /100

* Model 3 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_21_x100 a_wrkn_b_x100
	a_wrkn_c_x100 a_wrkn_d_x100 a_wrkn_e_x100 a_wrkn_f_x100
	a_wrkn_g_x100 a_wrkn_h_x100 a_wrkn_i_x100 a_wrkn_j_x100
	a_wrkn_k_x100 a_wrkn_l_x100 a_wrkn_m_x100 a_wrkn_n_x100
```

```
        a_wrkn_o_x100       a_wrkn_p_x100       a_wrkn_q_x100       a_wrkn_r_x100
        a_wrkn_s_x100       a_wrkn_u_x100, vce(cluster wk_code)
outreg2 using regmodels_21_sectors.doc, append ctitle (model 3)

* Model 4 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_21_x100             a_wrkn_b_x100
        a_wrkn_c_x100       a_wrkn_d_x100       a_wrkn_e_x100       a_wrkn_f_x100
        a_wrkn_g_x100       a_wrkn_h_x100       a_wrkn_i_x100       a_wrkn_j_x100
        a_wrkn_k_x100       a_wrkn_l_x100       a_wrkn_m_x100       a_wrkn_n_x100
        a_wrkn_o_x100       a_wrkn_p_x100       a_wrkn_q_x100       a_wrkn_r_x100
        a_wrkn_s_x100       a_wrkn_u_x100       woningen     avg_opp
        c_51_75     c_76_100     c_101_150    c_151_250    c_251_grot
        c_pre1900 c_1900_1944     c_1945_1979 c_1980_1999          c_unknown
        p_mgezw     p_koopwon             p_leegsw, vce(cluster wk_code)
outreg2 using regmodels_21_sectors.doc, append ctitle (model 4)

* Model 5 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_21_x100             a_wrkn_b_x100
        a_wrkn_c_x100       a_wrkn_d_x100       a_wrkn_e_x100       a_wrkn_f_x100
        a_wrkn_g_x100       a_wrkn_h_x100       a_wrkn_i_x100       a_wrkn_j_x100
        a_wrkn_k_x100       a_wrkn_l_x100       a_wrkn_m_x100       a_wrkn_n_x100
        a_wrkn_o_x100       a_wrkn_p_x100       a_wrkn_q_x100       a_wrkn_r_x100
        a_wrkn_s_x100       a_wrkn_u_x100       woningen     avg_opp
        c_51_75     c_76_100     c_101_150    c_151_250    c_251_grot   c_pre1900
        c_1900_1944 c_1945_1979 c_1980_1999          c_unknown    p_mgezw
        p_koopwon           p_leegsw            p_opp_water  af_artspr    af_ziek_i
        af_ziek_e    af_superm    af_warenh    af_cafe      af_restau    af_hotel
        af_kdv af_ondbas    af_ondvrt           af_brandw    af_oprith    af_treinst
        af_overst    af_zwemb     af_ijsbaan   af_biblio    af_zonbnk    af_attrac
        af_museum, vce(cluster wk_code)
outreg2 using regmodels_21_sectors.doc, append ctitle (model 5)

* Model 6 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100          a_wrkn_b_x100       a_wrkn_c_x100
        a_wrkn_d_x100       a_wrkn_e_x100       a_wrkn_f_x100       a_wrkn_g_x100
        a_wrkn_h_x100       a_wrkn_i_x100       a_wrkn_j_x100       a_wrkn_k_x100
        a_wrkn_l_x100       a_wrkn_m_x100       a_wrkn_n_x100       a_wrkn_o_x100
        a_wrkn_p_x100       a_wrkn_q_x100       a_wrkn_r_x100       a_wrkn_s_x100
        a_wrkn_u_x100       woningen     avg_opp                   c_51_75
        c_76_100     c_101_150    c_151_250    c_251_grot   c_pre1900    c_1900_1944
        c_1945_1979 c_1980_1999          c_unknown    p_mgezw      p_koopwon
        p_leegsw     p_opp_water  af_artspr    af_ziek_i    af_ziek_e    af_superm
        af_warenh    af_cafe      af_restau    af_hotel     af_kdv af_ondbas
        af_ondvrt           af_brandw    af_oprith    af_treinst   af_overst
        af_zwemb     af_ijsbaan   af_biblio    af_zonbnk    af_attrac    bev_dichth
        oad    sted              p_west_al    p_n_w_al     p_laaginkh   p_hooginkh
        p_socminh    p_nietact    ao_uit_tot   a_soz_ow     g_wodief     g_vernoo
        g_gewsek, vce(cluster wk_code)
outreg2 using regmodels_21_sectors.doc, append ctitle (model 6)

*** 9       Robustness test 3: testing non-linearity in the employment sectors (incl. creating
table + figures)
** generating squared variables
generate a_wrkn_sq_a = a_wrkn_a ^2
generate a_wrkn_sq_bcdef = a_wrkn_bcdef ^2
```

```
generate a_wrkn_sq_gi = a_wrkn_gi ^2
generate a_wrkn_sq_hj = a_wrkn_hj ^2
generate a_wrkn_sq_kl = a_wrkn_kl ^2
generate a_wrkn_sq_mn = a_wrkn_mn ^2
generate a_wrkn_sq_opq = a_wrkn_opq ^2
generate a_wrkn_sq_rstu = a_wrkn_rstu ^2

* Model 3 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
a_wrkn_sq_a a_wrkn_sq_bcdef a_wrkn_sq_gi a_wrkn_sq_hj a_wrkn_sq_kl
	a_wrkn_sq_mn	a_wrkn_sq_opq	a_wrkn_sq_rstu, vce(cluster wk_code)
outreg2 using regmodels_quadratic.doc, append ctitle (model 3)

* Model 4 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	a_wrkn_sq_a a_wrkn_sq_bcdef	a_wrkn_sq_gi a_wrkn_sq_hj a_wrkn_sq_kl
	a_wrkn_sq_mn	a_wrkn_sq_opq	a_wrkn_sq_rstu	woningen
	avg_opp		c_51_75	c_76_100	c_101_150	c_151_250
	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979 c_1980_1999
	c_unknown	p_mgezw	p_koopwon	p_leegsw, vce(cluster wk_code)
outreg2 using regmodels_quadratic.doc, append ctitle (model 4)

* Model 5 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	a_wrkn_sq_a a_wrkn_sq_bcdef	a_wrkn_sq_gi a_wrkn_sq_hj a_wrkn_sq_kl
	a_wrkn_sq_mn	a_wrkn_sq_opq	a_wrkn_sq_rstu	woningen
	avg_opp		c_51_75	c_76_100	c_101_150	c_151_250
	c_251_grot	c_pre1900	c_1900_1944 c_1945_1979 c_1980_1999
	c_unknown	p_mgezw		p_koopwon	p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	af_museum, vce(cluster wk_code)
outreg2 using regmodels_quadratic.doc, append ctitle (model 5)

* Model 6 - ln(P_n) = a + B1Dn + B2Dwa_n + B3X_n + B4Y_n + B5Z_n + B6mu_n + e_n
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100
a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100
	a_wrkn_sq_a a_wrkn_sq_bcdef	a_wrkn_sq_gi a_wrkn_sq_hj a_wrkn_sq_kl
	a_wrkn_sq_mn	a_wrkn_sq_opq	a_wrkn_sq_rstu	woningen
	avg_opp		c_51_75	c_76_100	c_101_150	c_151_250
	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979 c_1980_1999
	c_unknown	p_mgezw	p_koopwon	p_leegsw	p_opp_water
	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh	af_cafe
	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt	af_brandw
	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan	af_biblio
	af_zonbnk	af_attrac	bev_dichth	oad	sted		p_west_al
	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact	ao_uit_tot
	a_soz_ow	g_wodief	g_vernoo	g_gewsek, vce(cluster wk_code)
outreg2 using regmodels_quadratic.doc, append ctitle (model 6)

** Graphs
```

generate ln_woz_opq_sq = (((exp(-.0000889)-1)*100)*a_wrkn_opq) + (((exp(3.36e-08)-1)*100)*a_wrkn_sq_opq)
twoway (line ln_woz_opq_sq a_wrkn_opq, sort)

generate ln_woz_rstu_sq = (((exp(.0004298)-1)*100)*a_wrkn_rstu) + (((exp(-3.71e-07)-1)*100)*a_wrkn_sq_rstu)
twoway (line ln_woz_rstu_sq a_wrkn_rstu, sort)

*** 10 Robustness test 4: performing a Chow-test on population density (incl. creating table)
*Pooled model
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100 woningen avg_opp c_51_75 c_76_100 c_101_150 c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979 c_1980_1999 c_unknown p_mgezw p_koopwon p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e af_superm af_warenh af_cafe af_restau af_hotel af_kdv af_ondbas af_ondvrt af_brandw af_oprith af_treinst af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac bev_dichth oad sted p_west_al p_n_w_al p_laaginkh p_hooginkh p_socminh p_nietact ao_uit_tot a_soz_ow g_wodief g_vernoo g_gewsek, vce(cluster wk_code)
outreg2 using regmodels_Chow_bevdichth_3.doc, replace ctitle (Pooled)

*Group 1 - low population density (variable: bev_dichth)
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100 woningen avg_opp c_51_75 c_76_100 c_101_150 c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979 c_1980_1999 c_unknown p_mgezw p_koopwon p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e af_superm af_warenh af_cafe af_restau af_hotel af_kdv af_ondbas af_ondvrt af_brandw af_oprith af_treinst af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac bev_dichth oad sted p_west_al p_n_w_al p_laaginkh p_hooginkh p_socminh p_nietact ao_uit_tot a_soz_ow g_wodief g_vernoo g_gewsek if c_bev_dichth_3==1
outreg2 using regmodels_Chow_bevdichth_3.doc, append ctitle (Group 1)
*Group 2 - middle population density (variable: bev_dichth)
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100 a_wrkn_hj_x100 a_wrkn_kl_x100 a_wrkn_mn_x100 a_wrkn_opq_x100 a_wrkn_rstu_x100 woningen avg_opp c_51_75 c_76_100 c_101_150 c_151_250 c_251_grot c_pre1900 c_1900_1944 c_1945_1979 c_1980_1999 c_unknown p_mgezw p_koopwon p_leegsw p_opp_water af_artspr af_ziek_i af_ziek_e af_superm af_warenh af_cafe af_restau af_hotel af_kdv af_ondbas af_ondvrt af_brandw af_oprith af_treinst af_overst af_zwemb af_ijsbaan af_biblio af_zonbnk af_attrac bev_dichth oad sted p_west_al p_n_w_al p_laaginkh p_hooginkh p_socminh p_nietact ao_uit_tot a_soz_ow g_wodief g_vernoo g_gewsek if c_bev_dichth_3==2
outreg2 using regmodels_Chow_bevdichth_3.doc, append ctitle (Group 2)
*Group 3 - high population density (variable: bev_dichth)
reg ln_woz di_8 avg_di_8 a_wrkn_a_x100 a_wrkn_bcdef_x100 a_wrkn_gi_x100

```
a_wrkn_hj_x100  a_wrkn_kl_x100  a_wrkn_mn_x100  a_wrkn_opq_x100  a_wrkn_rstu_x100
	woningen	avg_opp			c_51_75	c_76_100	c_101_150
	c_151_250	c_251_grot	c_pre1900 c_1900_1944	c_1945_1979
	c_1980_1999		c_unknown	p_mgezw	p_koopwon		p_leegsw
	p_opp_water	af_artspr	af_ziek_i	af_ziek_e	af_superm	af_warenh
	af_cafe	af_restau	af_hotel	af_kdv af_ondbas	af_ondvrt
	af_brandw	af_oprith	af_treinst	af_overst	af_zwemb	af_ijsbaan
	af_biblio	af_zonbnk	af_attrac	bev_dichth	oad	sted
	p_west_al	p_n_w_al	p_laaginkh	p_hooginkh	p_socminh	p_nietact
	ao_uit_tot	a_soz_ow	g_wodief	g_vernoo	g_gewsek		if
c_bev_dichth_3==3
outreg2 using regmodels_Chow_bevdichth_3.doc, append ctitle (Group 3)
```