The influence of address density on livability and the relation of poverty to both

Course:	**Thesis Master Economics & Business**
Master:	**Urban, Port and Transport Economics**
Supervisor:	**Mr. M.J. Burger**
Number of pages:	**44**
Number of words:	**9860 (excluding front page, references & appendix)**
Date:	**20-07-2021**
University:	**ERASMUS University Rotterdam**
Faculty:	**Erasmus School of Economics**

Student Nummer	**Achternaam**	**Voornaam**
484198	Muntendam	Matthijs

There is a contradiction between livability and poverty on neighbourhood level. People want to live in dense areas, while these areas are simultaneously associated with higher shares of poverty. Previous research shows that the accessibility of a central marketplace is crucial for the value of land and that public transport amenities are the main reason for the sorting process of poverty. In this paper an Ordered Logistic regression is used to clarify the association of address density with livability. Furthermore, Ordinary Least Squares regressions are performed to test whether a higher address density is associated with more households with an income below the social minimum. The results show that a higher address density on the one hand is associated with a lower livability and on the other hand with higher shares of poverty in The Netherlands in 2018. The role of criminality might give an explanation for the obtained results.

Table of Contents

1. Introduction

The large municipalities in the Netherlands want to be more sustainable in the future. This would mean less cars, more skyscrapers, and an impulse for the green space. These developments happen simultaneously with the ending era of using farmland to build new neighborhoods. Using farmland to develop neighborhoods has turned out to be cheap for developers, but expensive for the society in which farmland is scarce. Besides, the green space and farmland in the Netherlands are important from both a climate and ecological perspective. Furthermore, living in cities is efficient and sustainable. Therefore, Amsterdam and Utrecht have plans to build respectively 150 thousand and 60 thousand houses within the city. Building in such quantities within the urban area results in an increase in density and asks for a clear urban plan (Dinther, 2021). So, if the intensions to change are strong and not just holding for Amsterdam and Utrecht, it might be important to investigate the relation between address density and livability in neighborhoods.

Simultaneously, the majors of the fifteen largest Dutch municipalities have asked for a structural half billion euros per year to tackle problems in vulnerable neighborhoods. The majors state that the Covid-19 crisis increased the poverty and health problems of 1 million inhabitants in sixteen extreme vulnerable neighborhoods. Even if, the government investment is large, the majors have faith that at the end the gains exceed the costs. Considering, more employed people, increasing housing prices, and a higher overall livability (Trouw, 2021).

However, livability in general has a different meaning for different people. An inhabitant of Den Haag in Zuid-Holland will see a good public transport connection as a sign of high livability, while an inhabitant of Sneek in Friesland will mention the availability of open waters for recreation as a driver of livability. This example represents the variety of conceptual frameworks in the society. In the literature a similar pattern is visible. Van Kamp, Leidelmeijer, Marsman & De Hollander (2003) recognized that it is important to construct a uniform, multidisciplinary conceptual framework to accumulate knowledge. The livability indicator of The Netherlands consists of five different dimensions which have different weights in the end score. On a country level this model developed by The ministry of Interior Affairs and Kingdom relations (2018) has a strong correlation with the different neighbourhood satisfaction scores given by inhabitants. In this thesis livability is defined as the degree to which the environment meets the requirements and wishes set by inhabitants.

According to Von Thünen (1830) high satisfaction scores lead to high demand and low satisfaction scores lead to low demand, which results in respectively dense and rural neighborhoods. Surprisingly, nowadays dense areas are associated with poverty, while the demand for these areas is high. There is a contradiction between on the one hand the number of people which have the desire to live in cities, and on the other hand increasing levels of poverty and inequality when the address density is higher.

In America poverty is centralized in dense areas and has been assigned to political choices and public infrastructure (Glaeser, Kahn & Rappaport, 2008). The Netherlands represents a similar geographic picture with significant issues related to poverty in the large cities (Het Parool, 2008). In theory, cities are likely to have a higher livability considering public transport amenities and scale economies, but in practice they often do not have a higher livability. There is more nuisance and insecurity in dense areas compared to rural areas (Leidelmeijer, Middeldorp & Marlet, 2018). An explanation might be that good public infrastructure in cities attracts poor people and that higher concentrations of poverty increase the number of criminal incidents. Since, criminality is an important determinant of livability, higher shares of poverty are associated with lower livability levels.

This research investigates the relation between density and livability in the first place. Therefore, the Dutch government can reduce the housing shortage taking care of livability standards. Furthermore, it is important to clarify the link between dense areas and poverty to increase overall living standards by changing the structure of the neighborhoods. The results of this research can stimulate local governments to tackle the source of poverty to decrease high concentrations of poverty in neighborhoods. Moreover, an important determinant of livability is used to clarify the association between poverty and livability for different density levels. The results might help to take future decisions in de field of urban planning. Therefore, the research question is formulated as follows:

What is the influence of address density on livability and how is poverty related to both in the Netherlands in 2018?

The scientific relevance of this research is driven by the contribution to current papers which investigated livability and poverty. Previous research states that the availability of public transport plays a significant role in the centralization of poor people in dense areas (Glaeser et al., 2008). The research of Glaeser et al. (2008) investigates twelve metropolitan areas. Instead, my research has a wider range which contains all neighborhoods of the Netherlands, including rural areas in the north and east. The wider range gives the possibility to draw conclusions for the whole country and to use municipality fixed effects. Furthermore, my research controls the idea that poverty in neighborhoods lowers the overall quality of life (Hooimeijer & Van Kempen, 2000), via one of the determinants of livability. By doing regressions of criminality on poverty conditional on address density I use dependent variable which is likely to be affected by poverty in the short run, to test whether poverty has an effect on livability.

This paper has a theoretical framework in which three the hypotheses based on previous literature are formulated. The section data discusses which data variables are used and contains descriptive statistics. The method contains multiple ordered logistic regressions and several ordinary least squares regressions to test the three hypotheses. Furthermore, this section discusses the limitations and requirements for the used methods. Followed by the obtained results from the performed analysis in the section results. In the conclusion and discussion, the research question is evaluated and the results are placed in perspective.

2. Theoretical Framework

This section contains the theoretical framework to estimate what is the influence of address density on livability and how is poverty related to both in the Netherlands in 2018. The research question is answered by testing three hypotheses. Firstly, the expected influence of address density on the level of livability is tested. Secondly, the relation of address density with poverty is clarified by using research done by (Glaeser, Kahn & Rappaport, 2008). Thirdly, the association between poverty and livability is estimated by using criminality as important volatile determinant of livability. This structure allows to clarify the influence of address density on livability and to figure out the relation of poverty with both address density and livability.

2.1 Livability

Livability has a different meaning for different people and in the literature, there are many conceptual frameworks. Despite the variety of conceptual frameworks in the literature Van Kamp, I., Leidelmeijer, K., Marsman, G. & De Hollander, A. (2003) recognized that it is important to construct a uniform, multidisciplinary conceptual framework to accumulate knowledge. Their concept has three main directions, the physical direction, the economic direction, and the social direction. In addition, the same three determinants are used by Shafer, Lee & Turner (2000) in their quality-of-life concept. Furthermore, Camagni, Capello & Nijkamp (1997) used the three determinants to create an approach towards sustainability in cities. The framework of Van Kamp et al. (2003) allows to accumulate knowledge, which is important to know where to intervene as a government and what is the effect of government interventions on livability. Mainly driven by the advice of Kamp et al. (2003) The Dutch ministry of Interior Affairs and Kingdom relations introduced a so called ‘Leefbaarometer’.

The ‘Leefbaarometer’ focusses on the degree of livability on different spatial scale levels. Livability is described as the degree to which the environment meets the requirements and wishes set by inhabitants. The livability indicator for the Netherlands consists of five different dimensions, which in total contain 100 variables. The dimensions are houses, inhabitants, amenities, safety, and physical environment. The dimensions have different weights with the aim to have a low standard deviation. Amenities (healthcare, retail, catering industry, education, transport and recreation) have the most important role in the model with 25%, followed by safety of 24%, physical environment and houses represent both 18% and inhabitants have a relatively low share of 15% (The ministry of Interior Affairs and Kingdom relations, 2018).

Analyzing different variables to determine a low or a high livability is closely related to hedonic demand theory. Research into the demand for a certain product by splitting up different characteristics, also called hedonic demand theory, is a revealed preference method to estimate the demand and therefore the value of a product (Rosen, 1974). This makes it possible to give value to the different characteristics of a location. When location A has more favorable characteristics compared to location B, the utility of consumers is higher at location A than at location B. Therefore, demand increases for location A and decreases for location B. As a result, two identical buildings have a different level of demand according to the location. This results in different prices and rents for houses.

Research between the location of real estate and the value of the property has been done for more than 190 years. It turns out that there is a relation between the value of land for agriculture and the availability of a marketplace. Better connectivity results in lower transport costs, which makes it possible for individuals with lower benefits to obtain a positive utility. When there are more farmers with a positive utility, there is a higher aggregated demand for land with an agricultural function. In this way you can declare that land with the same fertility, but a different location differs in value. (Von Thünen, 1830).

Neighborhoods with a lower livability level have inhabitants which are less likely to walk, cycle or garden in leisure time. Furthermore, these inhabitants are less likely to participate in sport activities (Van Lenthe, Brug & Mackenbach, 2005). Therefore, they have a lower overall utility level. The lack of enjoying life and participating in sport activities results respectively in lower benefits and higher health insurance costs. The lower utility level leads to a lower aggregated demand for houses. On the contrary, inhabitants get a higher utility for living in places with a higher livability level, which leads to demand increases probably causing higher housing prices.

Considering that the supply side of houses is determined by a production relation, which is constant in the short run, an increase in demand leads directly to higher prices (Poterba, 1984). However, if housing supply is relatively elastic, an outward shift of demand results in an increase of the local population. Simultaneously, the increase in prices in case of elastic supply causes a rise in price above construction costs (Glaeser, Gyouko & Saks, 2008).

According to the theory of Von Thünen (1830) you can expect that nowadays transport amenities like railway and road infrastructure determine a part of livability. Furthermore, these amenities influence address density via the market demand. The combination of an increase in demand and a relative elastic housing supply leads by assuming fixed geographical borders to an increase in address density. These increase in address density might have some interesting consequences for the livability of a neighborhood considering the definition of a livable place made by Balsas (2010). He states that a livable place is safe, clean, beautiful, economically vital, affordable to a diverse population and efficiently administered, with functional infrastructure, interesting cultural activities and institutions, ample parks, effective public transportation, and broad opportunities for employment.

The level of economic vitality is linked to the number of viable businesses and profitable investments. Keeping this in mind, an increase in address density leads to more potential customers in a neighborhood and thus a higher demand for your products as entrepreneur. In the market were demand and supply meet this increase in demand leads to a higher chargeable price. So, a higher price means that more businesses are viable, and investment is more profitable. Concluding, a higher address density leads to a higher level of economic vitality.

The quality of amenities is two sided. On one hand a higher address density leads to more demand for all amenities, like cultural activities and institutions. On the other hand, higher address density can result in disamenities such as crime and congestion. However, I assume that in the Netherlands by the time most of these problems of dense areas are tackled and therefore the first hypothesis is as follows:

Hypothesis 1: Higher address density is associated with a higher livability.

2.2 Poverty

According to the first hypothesis, dense areas are expected to have a higher livability than rural areas. Strikingly, is the fact that the percentage of poor people in dense areas is higher than the percentage of poor people in rural areas (Margo, 1992, Mieszkowski & Mills, 1993, Mills & Lubuele, 1997). A central question in urban economics has been why poor people live disproportionately in dense areas. The land use theory of Alonso (1964) and Muth (1968) states that rich people demand relatively more land than poor people and therefore choose to live outside the city center where land is cheaper. The theory holds if the income elasticity of demand for land is larger than the income elasticity of travel cost per mile. However, using empirical evidence it is unlikely that the income elasticity of demand for land is larger than the income elasticity of travel cost per mile. Indeed, the income elasticity of demand is equal to 0.4 and income elasticity of travel costs is equal to one (Glaeser, Kahn & Rappaport, 2008).

On an aggregated level, the economic theory which explains why urbanization has economic benefits and what are these benefits has improved a lot (Glaeser, 2011; Krugman, 2011). There is a positive relation between urbanization, the proportion of the total national population living in areas classed as urban, and per capita income. Furthermore, evidence from high income countries shows that larger urban settlements have a higher productivity than the smaller ones (Turok and McGranahan, 2013). However, these theories do not explain the higher percentages of poverty in dense areas.

Instead, we know from simple economic theory that on the market more supply results in a lower equilibrium price. Therefore, you could argue that the higher the address density, the more housing supply, causing a higher share of the population which can afford to live in a specific place due to lower prices. Justified by utility theory, which says that lower housing costs make it affordable for a larger population share to reach a positive utility. This indicates that people with lower income benefits can make the decision to live in a dense neighborhood and would therefore sort in dense areas. However, Glaeser et al. (2008) find that the cost of housing of the poor relative to the rich do not decline in dense areas. Furthermore, it turns out that in dense areas high concentrations of poor people correlate with high incentives for poor people to live in less dense areas. Indicating that the structure of the housing market cannot explain centralized poverty.

The housing market itself cannot explain centralized poverty, but the distortion of the housing market by political choices can. The strong positive relationship between public housing and poverty indicates that places with a better public housing policy might be more attractive for poor people, supported by two reasons. Firstly, poor people have a nearly 10 percent bigger chance to use public housing facilities in a dense area. Secondly, the benefit of living in a subsidized housing area is bigger in dense areas than in it is in rural areas (Glaeser et al., 2008). Taking these two empirical facts into account you might expect that political choices regarding public housing influence the centralization of poverty into dense areas.

In addition to political choices the availability of public transport plays an important role in the sorting of poor people. Public transportation is a relatively cheap transport mode and in general better available in dense areas, because of more demand and historical heritage. Regressing median income on public transportation proximity fades away the relation between distance to the city center and income. Within the association between income and public transport reverse causality might be a problem. Did cities invest in public transport to serve the poor? However, most of the public transportation amenities have been built long ago, which would mean that any form of endogeneity stems from poverty levels in the past (Glaeser et al., 2008). The revolution of Harlem into a ghetto, started by the extension of an existing metro line, is an example which proves even stronger the relation between public transport in the direction of poverty (Osofsky, 1966).

A closer look into the model with three modes of transportation shows that in old cities rich people living very close to the center walk to work and rich people living in suburbs use the car to travel. In between poor people live, who use the public transport. Moreover, it turns out that in regions where just one mode of transportation is used, in this example the car, rich people tend to live in areas with a higher density. This means that the availability of several modes of transport is crucial to understand why poor live in dense areas (Glaeser et al., 2008). Furthermore, the fact that a decrease in distance of one kilometer to a railway station results in a 1.3% decrease of housing prices in Dutch neighborhoods runs parallel to the idea that poor sort around good public transport amenities (Muntendam, 2020). Considering that poor sort in dense areas because of more public housing and better public transport amenities the second hypothesis is as follows:

Hypothesis 2: Higher address density is associated with higher shares of poverty in a neighborhood.

2.3 Poverty and criminality conditional on address density

According to the two hypothesizes higher address density has a positive association with livability but leads also to higher shares of poverty. This is striking since it would mean that neighborhoods with a higher livability have a higher share of poverty, while mostly a negative relation is assumed. Probably one of the key determinants of livability, the number of criminal accidents in a neighborhood, plays an important role in the interaction between livability and poverty. The level of poverty in a neighborhood is strongly associated with the level of crime in a neighborhood (Patterson, 1991). Furthermore, I choose for the number of criminal incidents representing the determinant safety, because this determinant is likely to be affected by poverty in the short run. The total livability level is less because it contains also fixed determinants as housing. Intuitively, poverty might change the number of steals, but not the average number of bathrooms in a neighbourhood.

The centralization of poverty and therefore increasing crime rates can have a multiplier effect according to the paper of Cullen & Levitt (1999). They find that people with a higher education are more likely to move out of a crime center than people with a lower education. The same holds for households with children which are more able to migrate. This results in a lower average level of education in a neighborhood and more aging. In the Netherlands, a similar research found that for some groups in the population it is more difficult to move out of a neighborhood with high levels of poverty. Minority ethnic groups have the largest problem to move away from poverty areas. These groups have the desire to be located closely to their own group. Besides, the high concentration of poverty neighborhoods in dense areas does not ease the moving problem. Since, the large cities have many poverty neighborhoods from an absolute point of view, which makes better housing alternatives scarce (Bolt & Van Kempen, 2010).

Despite that, it must be mentioned in The Netherlands cleavages are small and that the conditions of neighborhoods have a little impact on the life chances of inhabitants (Bolt, Burgers & Van Kempen, 1998; Musterd, 2002). However, the level of livability is affected by poverty via criminality in Dutch neighborhoods. Wittebrood (2000) found that the risk of violent victimization increasing due to poverty and other research showed that poverty lowers overall quality of life (Hooimeijer & Van Kempen, 2000).

Furthermore, there are some scale economies in cities, which help cities usually to be more productive, increasing the return to criminal activities (Glaeser & Sacerdote, 1999). Moreover, they found again that these high crime rates mostly derive from high concentrations of poor people in dense areas and not from intrinsic issues in the areas themselves. Therefore, the third hypothesis is as follows:

Hypothesis 3: Poverty is stronger associated with criminal activities in dense areas.

3. Data

This section describes the used data. The first and second hypothesis are tested by using cross-sectional data obtained in 2018. All the variables in the regressions are explained. However, when a variable is included in both regressions it is explained just once. The third hypothesis is tested using cross-sectional data with three added criminality determinants.

3.1 Cross-sectional data

The cross-sectional data is used to test the first two hypotheses. In these data two dependent variables are included. The livability level for 10986 neighborhoods in 2018 (Leefbaarometer, 2018) and the number of households below the social minimum income level for 9112 different neighborhoods (CBS, 2019a). The independent variable address density per square kilometer, which is the main variable of interest for both hypotheses, has 13242 observations in the dataset.

3.1.1 Livability scores

Table 1: Frequency of livability levels in 2018.

Livability in 2018	Frequence	Percentage	Cumulative
1	13	0,12	0,12
2	24	0,22	0,34
3	91	0,83	1,17
4	427	3,89	5,05
5	521	4,74	9,79
6	2389	21,75	31,54
7	2796	25,45	56,99
8	2349	21,38	78,37
9	2376	21,63	100
Total	10986	100	

The livability indicator for the Netherlands consists of five different dimensions, which in total contain 100 variables. The dimensions are houses, inhabitants, amenities, safety, and physical environment. Combining the dimensions results in an ordinal dependent variable with 9 different categories. A neighborhood has a level equal to one if it has a very low livability and a level equal to 9 when it has a very high livability. From table 1 we obtain that first five categories represent just a small 10% of the data. Instead, the categories six till nine represent all between 21% and 26% of the data. This indicates that the Netherlands has many neighborhoods with a high livability and a relatively small number of neighborhoods with a low livability level. The same results are obtained from the histogram in figure 1 in the appendix, where we see a left tailed distribution. However, the small number of low livability level observations says nothing about the size of the problems within these neighborhoods.

3.1.2 Address Density

Table 2: Descriptive statistics of several variables in 2018.

Variable	Obs	Mean	Std. Dev.	Min	Max
Address Density	13 242	1183.4	1493.6	0	12389
Proximity Railway Station	12 682	6.5	6.9	0.2	60.3
Proximity Transfer Station	12 682	13.1	10.1	0.2	71.9
Proximity Main Road	12 682	1.9	1.8	0.1	46.4
Households Below Social Minimum	9 112	6	4.6	0	48.9
Average Housing Price	10 099	261.5	120.4	37	1836
Public Rental Properties	11 836	18.5	21.1	0	100

The variable address density gives the number of addresses within a square kilometer for 13242 neighborhoods. There is a neighborhood with 0 addresses, which is probably a natural area or a place which is under construction. The maximum addresses per square kilometer is equal to 12389, much higher than the mean of 1183. The histogram in figure 2 in the appendix shows that the address density in 2018 is skewed to the right. Indicating that the Netherlands has a lot of neighborhoods with a low density and few with a very high density. This characterizes the geographic structure of the country, some large cities with dense urban areas, surrounded by a lot of little villages with low density scores.

The correlation between address density and livability level is negative in The Netherlands in 2018 (table 3 in the appendix).

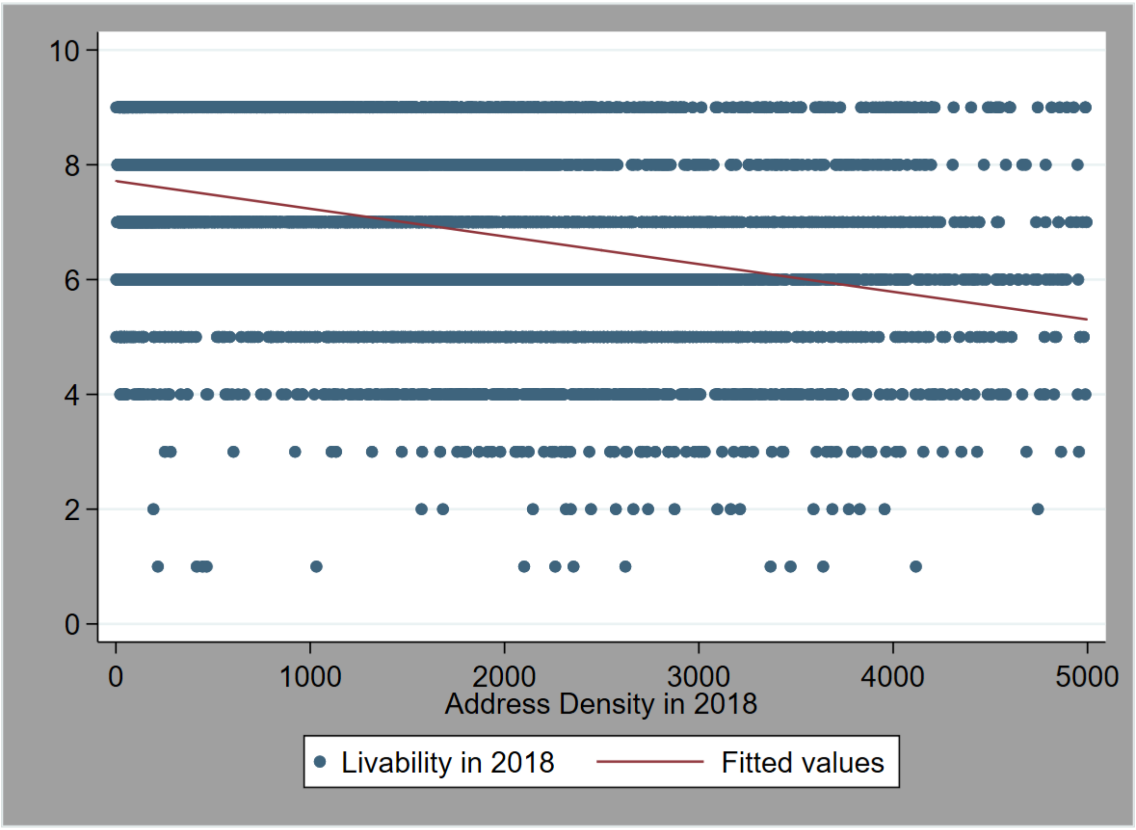

Figure 3: Scatterplot of livability plotted against address density.

Figure 3 seems to show the same negative relation between address density and livability. However, this is just a partial correlation if there is not controlled for characteristics of neighborhoods which both influence address density and livability. Furthermore, the ordinal structure of the dependent variable is visible because we have nine horizontal lines. This results in an error term with a non-normal distribution, unbounded predicted probabilities and heteroskedasticity when a linear probability model is performed. Therefore, I use an ordinal logistic regression, which is explained in the section methods.

3.1.3 Proximity variables

In addition, we need several proximity variables to reduce Omitted Variable Bias. Omitted Variable Bias occurs when the independent variable is not completely exogenous. This is further explained in the section methods. From the literature it turns out that variables which represent connectivity influence both livability and address density. The variables proximity of the railway station, proximity of the closest transfer station, and proximity of a main road are available for 12682 neighborhoods in the Netherlands (CBS, 2019b). The variables are measured in kilometers, accurately on one decimal. The minimum distance to a railway station is 0.2 kilometer, while the maximum is equal to 60.2 km. The country average on neighborhood level is 6.5 kilometer.

The average distance to a transfer station is slightly higher, namely 13.1 kilometers. The minimum distance towards a transfer station is again 0.2 kilometers. This shows that there are neighborhoods with an important transfer station as closest railway station, for example Utrecht Central Station.

The average proximity of a main road is with 1.9 kilometers of distance significantly lower than the variables related to public infrastructure. An inhabitant of the neighbourhood with the longest travel towards a main road must travel 46.4 kilometers. On the other side the minimum travel distance is equal to 0.1 kilometer, again lower than the variables related to public transport.

3.1.4 Households below the social minimum income level

The second dependent variable is the percentage of households below the social minimum income level. The social minimum income is equal to €1079,-- per month for a single person in the Netherlands. For this variable we have 9112 observations, lower than the proximity variables. The main reason is that in the previous variables business parks are classified as neighbourhood, but it makes sense that households are not located on a business park. The average percentage of households below poverty level is equal to 6.0 %. The highest percentage of households below the social minimum is a neighbourhood with almost half of the households in this category, namely 48.9%. On the contrary, the minimum level of poverty within a neighbourhood is equal to zero percent. From figure 4 in the appendix, it can be obtained that the distribution of the poverty variable is skewed to the right. In other words, relatively a lot of neighborhoods have a small percentage of poverty households, and a few neighborhoods have high percentages of poverty. This is line with the picture for the livability indicator.

The correlation between address density and households below the social minimum income level is positive (table 3 in the appendix).

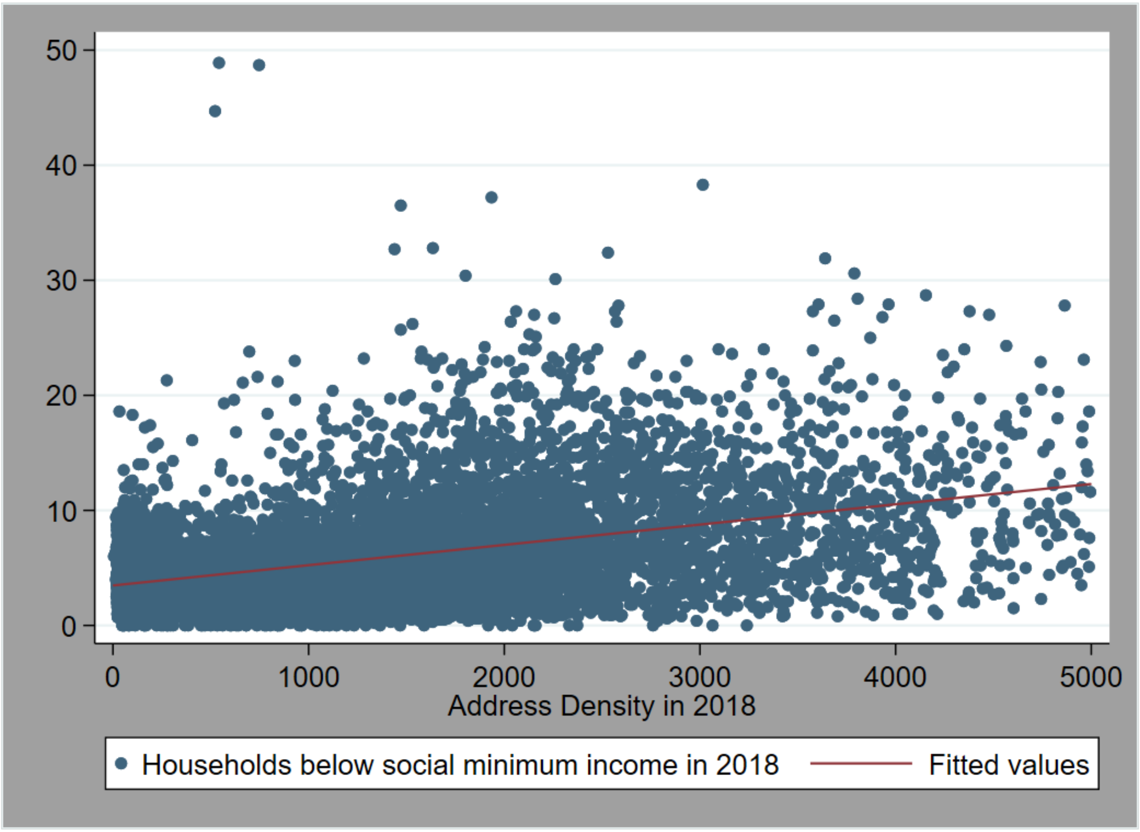

Figure 5: Scatterplot of households below social minimum income plotted against address density.

Figure 5 shows the same positive relation between address density and households below the social minimum in 2018. Even if this relation is similar to the ones described in the literature it is just a partial correlation as long as I do not control for other neighborhood characteristics which influence both variables. Furthermore, the scatterplot shows that not all the actual values for poverty have the same distance towards the fitted values for every level of density. This is an indication of heteroskedasticity, which is further explained in the section methods.

3.1.5 Neighbourhood characteristics

The average housing price, the percentage of social rent properties and again the proximity of the railway station are important to reduce Omitted Variable Bias. The average housing price in 2018 in the Netherlands is equal to €261 523,--. The neighborhoods with respectively the lowest and highest housing prices have an average of €37 000,-- and 1.8 million euros. This shows that the differences are big.

The percentage of rental properties is available for 11836 different neighborhoods, slightly more than the average housing price. On average, 18.5 percent of the total market for housing properties has the social status. The differential between the minimum of the maximum value is exactly 100. Meaning that, in the Netherlands there are neighborhoods with zero and all houses available for social rent.

3.1.6 Fixed effects

Finally, municipality fixed effects are used to control for all time-irrelevant factors, which influence poverty and address density on municipality level. Every neighbourhood in the dataset is linked to a certain municipality to use fixed effects. The Netherlands had 380 municipalities in 2018, which means that 379 dummies are included in the regressions.

3.2 Cross-sectional data with criminality

This section contains the description of data to test the third hypothesis. The dataset contains the number of criminal accidents per 1000 inhabitants. Criminality is one of the most important determinants of the livability level in a neighborhood. Previous literature showed that criminality is caused by poverty.

The number of criminal incidents per 1000 inhabitants is calculated as follows:

$$Criminal\ incidents_{i,j,y} = Thefts_{i,j,y} + Destruction_{i,j,y} + Violence_{i,j,y}$$

In this formula $Thefts_{i,j,y}$ represents the number of Thefts within real estates, like houses, sheds, and offices. Furthermore, $Destruction_{i,j,y}$ is the number of times that there has been destruction of the public space, for example of trash cans or playground equipment. Moreover, $Violence_{i,j,y}$ refers to behavior with violence, for example the number of rapes or violence against the police.

The variable is calculated on a neighborhood level for 2018. The neighborhood with the maximum number of criminal incidents has almost 2 incidents per inhabitant, namely 1953 per 1000 inhabitants. However, there are also neighborhoods with zero incidents and the average is with 15 incidents per 1000 inhabitants much lower than the maximum. Furthermore, previous discussed variables like percentage of households below the social minimum income level and address density per square kilometer are used to test the third hypothesis.

4. Methods

This section contains the used methods to test the three hypotheses. Firstly, an ordered logistic regression is chosen to test the hypothesis regarding livability. Secondly, an Ordinary Least Squares regression and the related assumptions are presented to test the hypothesis about poverty. Finally, several quartile regressions are used to find out the relation between criminality and poverty conditional on address density.

4.1 Livability

The first hypothesis is tested by using a quantitative research method. An ordered logistic regression is performed to test if a higher address density leads to a higher livability. The ordinal dependent variable $Livability_{i,j,2018}$ can take nine different values listed from 1 till 9:

$$Livability_{i,j,2018} \in \{1 < 2 < 3 < \cdots < 9\}$$

This means that the ordered logistic model has $k-1$, so 8 threshold points and that the cumulative probability of being in category 9 or lower is equal to 1. The model estimates are calculated by using several continuous variables at neighborhood level. The 8 multiple ordered logistic regression models are depicted as follows:

$$Cumulative\ \text{Logit}\ (Livability_{i,j,2018} \leq k) = a_k - (B_1 Address\ Density_{i,j,2018} + B_2 Prx\ Railway\ Station_{i,j,2018} + B_3 Prx\ Transfer\ Station_{i,j,2018} + B_4 Prx\ Main\ Road_{i,j,2018} + \varepsilon_{i,j,2018})$$

In this regression $Address\ Density_{i,j,2018}$ is the independent variable for each neighborhood i in municipality j in 2018. The dependent variable $Cumulative\ \text{Logit}\ (Livability_{i,j,2018} \leq k)$ is the cumulative logit value for each livability level. This means that the coefficient B_1 gives the change in cumulative logit value if there is one additional address per square kilometer. The model estimates are subtracted from the different thresholds a_k to calculate the cumulative logit value of being under or at a specific threshold. To get cumulative probabilities the following formula is performed:

$$Cumulative\ \text{Probability}\ (Livability_{i,j,2018} \leq k) = \frac{e^{Cumulative\ \text{Logit}\ (Livability_{i,j,2018} \leq k)}}{1 + e^{Cumulative\ \text{Logit}\ (Livability_{i,j,2018} \leq k)}}$$

Furthermore, to calculate the probability of ending up within a specific livability level, the cumulative probabilities are subtracted from each other. For example, $Cumulative \text{ Pr } (Livability_{i,j,2018} \leq 7) - Cumulative \text{ Pr } (Livability_{i,j,2018} \leq 6)$ gives the probability of $Livability_{i,j,2018}$ being equal to 6. However, to find the overall effect of a higher address density on the level of livability, average marginal effects are used. This allows to obtain the increase or decrease in the probability when the address density increases by one for every livability level.

Since I am interested in the causal effect of address density on livability the regression model is extended to meet the Zero Conditional Mean Assumption. Zero Conditional Mean assumes that the error term ε is independent from the address density, the independent variable. The expected value of the error term needs to be the same regardless of the number of addresses per square kilometer:

$$E\left(\varepsilon_{i,j,2018} \middle| Address\ Density_{i,j,2018}\right) = 0$$

When de *Zero Conditional Mean* assumption is fulfilled, the following holds:

$$corr\left(\varepsilon_{i,j,2018}, Address\ Density_{i,j,2018}\right) = 0$$

Since we obtain from the literature that the independent variable is not completely exogenous, several correlations are tested in table 5 in the appendix. At neighborhood level variables are used which both influence the address density and the level of livability. According to the literature variables which explain connectivity are important. Proximity of the railway station and proximity of the closest transfer station show both a negative correlation with address density. Indicating that if a neighborhood is closer to the public transport amenities it has also a higher address density. The correlation between proximity of the closest main road is and address density equal to zero. A possible declaration is that the Dutch road infrastructure is strongly developed, wherefore every neighborhood has a close main road, regardless the number of addresses.

Even though, the variables above prevent for Omitted Variable Bias. It is still difficult to meet the Zero Conditional Independence assumption completely. The assumption demands that conditional on all control variables the independent variable must be random. I cannot guarantee the complete randomness since it is not said that all control variables are covered in the dataset. However, the number of factors which influences the address density of a neighborhood is relatively low.

The control variables need to be determined before the independent variable. Instead, when variables are influenced by the address density, possible mechanisms or colliders are added into the regression. A mechanism is a variable with predictive power for the address density, but which also predicts the level of livability. A collider is determined by both address density and livability. Mechanisms and colliders introduce respectively the problem of selection bias and the problem of observing spurious relations. Therefore, mechanisms and colliders should not be included.

The first added control variable is the proximity of the closest railway station. From previous research it turns out that shorter distance to the railway station leads to a higher livability. Therefore, I expect the coefficient B_2 to be negative. Furthermore, the negative correlation between address density and proximity of public transport is supported by the theory of Von Thünen (1830) about the availability of the marketplace. The negative coefficient and the negative correlation probably show that the previous estimated coefficient B_1 is higher than the actual coefficient. For the average marginal effects this means that the probability of living in a neighbourhood with a low livability level increases when the address density is higher.

The second added control variable is the proximity of the closest transfer station. The variable is highly correlated with the proximity of the railway station, because when the distance to the railway station increases, the distance to the transfer station increases too. Despite that, I keep the two variables separated in the regression, because in rural areas distance to the transfer station might have a higher value than in dense areas. For the fact that the proximity to the closest transfer station is highly correlated with proximity of the railway station, coefficient B_3 is again expected to be negative. The previous estimated coefficient B_1 is might have an upward bias, because of the negative correlation between address density and proximity of the transfer station. In addition, a negative coefficient and correlation is expected for the third control variable the proximity of a main road, coefficient B_4. However, it must be mentioned that there is no correlation between address density and main road. Despite that it might be an important control variable because it is seen as a substitute for public amenities.

4.2 Poverty

The second hypothesis is tested with a quantitative research method. An ordinary Least Squares regression is performed to test whether address density has a positive relation with the percentage of households below the social minimum income. The dependent variable is a percentage, while the independent variable is natural number. Therefore, the multiple linear regression looks as follows:

$$Households\ below\ Social\ Minimum\ Income_{\,i,j,2018} = a + \ B_1 Adress\ Density_{i,j,2018} + \\ B_2 Average\ Housing\ Price_{i,j,2018} + B_3 Social\ Rent\ Properties_{i,j,2018} + \\ B_4 Prx\ Railway\ Station_{i,j,2018} + \sum_{j=2}^{380} \gamma Mun_j + \varepsilon_{i,j,2018}$$

In this regression $Adress\ Density_{i,j,2018}$ is density of addresses per square kilometer in neighborhood i within municipality j in 2018. The variable $Households\ below\ Social\ Minimum\ Income_{i,j,2018}$ is the dependent variable on neighborhood level in percentages. This indicates that the coefficient B_1 shows the change in percent point of the dependent when the address density increases by 1.

The use of an Ordinary Least Squares regression has some implications for the number of assumptions. This method has more assumptions than the logistic regression. I discuss some important ones. Firstly, the sample must have a unique and independent distribution, in other words no autocorrelation. This means that the error terms of the observations are not correlated:

$$Cov\left(\varepsilon_1, \varepsilon_2 \middle| Address\ Density_{i,j,2018}\right) = 0\ for\ 1 \neq 2$$

The used dataset has a cross-sectional format, this implies that observations do not have autocorrelation with each other. All the data variables are obtained in 2018, which means that the dataset does not contain time-series variables. Therefore, the assumption that errors are independent and uniquely distributed fits.

The second assumption is referring to homoskedasticity, in other words all error term must have the same variance conditional on the dependent variable:

$$Var\left(\varepsilon \middle| Address\ Density_{i,j,2018}\right) = \ \sigma^2$$

Figure 5 in the section data shows that not for every level of address density the actual values have the same distance to the fitted values. This implies heteroskedasticity. The fact that there is no constant variance of the error term is solved by using heteroskedastic-robust standard errors.

The third assumption is about multicollinearity. This condition states that it is not allowed to have a perfect relation between independent variables in the regression. Table 5 in the appendix shows that none of the independent variables has a strong correlation. The strongest negative correlation appears between housing prices and average public rental properties. This indicates that neighborhoods with higher housing prices are associated with less public rental properties.

The Zero Conditional Independence assumption has been widely discussed for the ordered logistic regression. Instead, now I search for variables which both influence address density as the percentage of households below the social minimum income level. The aim of satisfying this last assumption is to decrease Omitted Variable Bias, to estimate a causal relation.

The first variable added is the average housing price in 2018. For buying properties with higher values usually more income is needed, which increases the average earned income in a neighbourhood. Therefore, the average housing price is likely to have a negative coefficient B_2 with the percentage of households below the social minimum income level. Moreover, it is likely that higher housing prices indicate more square meters per house, and thus less addresses. This results in a negative expected correlation. Therefore, I expect that the previous estimated coefficient B_1 is higher than the actual coefficient, due to the negative coefficient and the negative correlation. However, the correlation of average housing price with address density is ambiguous in table 5 in the appendix. Average housing price might be a mechanism when the historic address density determines the housing price today. Mechanisms introduce selection bias, so I need to be careful with including average housing price into the regression.

The second variable added is the percentage of public housing per neighborhood. Since public housing is a policy for lower income groups the coefficient B_3 is likely to be positive. More public housing available leads to more households with a low income, and thus also the chance that they live below social minimum. Furthermore, when a partial or whole neighborhood is classified as public housing area the address density will be higher. Housing corporations usually develop smaller houses and houses closer to each other than private companies do. This makes the correlation between percentage of public housing and address density positive. The positive coefficient and correlation suggest that the previous estimated coefficient B_1 is higher than the actual coefficient.

The third variable is added since the literature shows that the availability of public transport amenities attracts poor people. Therefore, the proximity of a railway station is expected to have a negative coefficient B_4. When the distance to the station decreases the percentage of poverty households increases. Moreover, the correlation between proximity of the railway station and address density is negative. Intuitive, better connectivity results in higher demand and thus more addresses per square kilometer. The negative coefficient and negative correlation probably show that the previous estimated coefficient has an upward bias.

The last controls which are added are municipality fixed effects. The Netherlands has 380 municipalities in 2018. These fixed effects control for time-irrelevant differences relative to one single municipality. This means that in total 379 dummies are added to control for differences between municipalities, which both influence address density as also the percentage of households below the social minimum income.

4.3 Poverty and criminality conditional on address density

The third hypothesis is tested by using a quantitative research method, using the same cross-sectional data as before with added criminality indicators. Criminality is one of the main determinants of livability and therefore interesting to review in combination with the percentage of households below the social minimum income level. From the literature two main insights appear. Firstly, higher levels of poverty are associated with more criminal incidents. Second, scale economies in cities make criminal activities more productive.

In line with what is written above I perform an Ordinary Least Squares regression to see whether more households living below the social minimum income level are associated with more criminal incidents in Dutch neighborhoods in 2018. Criminal incidents are determined as stated in the previous section data. To test if cities make criminal activities more productive, I include the number of addresses per square kilometer into the regression. Besides, adding the interaction term between percentage of households below the social minimum and address density allows to measure if economies of scale in dense areas lead to a stronger association between poverty and criminality.

5. Results

This section contains the results for the three hypotheses. Firstly, several Ordinal Logistic regressions are performed to test whether a higher address density leads to a higher livability. For each model, the average marginal effects of address density are calculated. Secondly, Ordinary Least Squares regressions are estimated to test if a higher address density leads to a higher percentage of households having a lower social minimum income. The different regressions have single and multiple formats since several control variables are added. The controls have been added based on previous literature and the assumptions of Ordinary Least Squares to solve Omitted Variable Bias. Thirdly, several regressions conditional on address density are performed to see how criminality and poverty interact with different levels of address density.

5.1 Livability

Table 6: Summary of ordinal logistic regressions with the address density as determinant of livability level in 2018.

	(1) Livability		(2) Livability		(3) Livability		(4) Livability	
Address Density	-0,00035	***	-0,00045	***	-0,00055	***	-0,00055	***
Railway Station			-0,05129	***	-0,02105	***	-0,02091	***
Transfer Station					-0,04138	***	-0,04132	***
Main Road							-0,00387	
Threshold 1	-7,35036		-7,89992		-8,4383		-8,44439	
Threshold 2	-6,30008		-6,8492		-7,38717		-7,39324	
Threshold 3	-5,04086		-5,58791		-6,12413		-6,13015	
Threshold 4	-3,4959		-4,03243		-4,5576		-4,56313	
Threshold 5	-2,74766		-3,27469		-3,79176		-3,79695	
Threshold 6	-1,21108		-1,69618		-2,18775		-2,19272	
Threshold 7	-0,09632		-0,54802		-1,01249		-1,01751	
Threshold 8	0,934545		0,506911		0,070001		0,064876	

* p<0.10, ** p<0.05, *** p<0.01

Regressing the independent variable address density on the dependent livability indicator results in a negative coefficient of -0.00035 significant on a 1% level for address density in table 6. This coefficient tells us that when the address density in a specific neighbourhood increases with 1000 addresses per square kilometer the cumulative logit increases with 0.35. A neighborhood with thousand addresses per square kilometer more has lots of high-rise architecture built in 1970, compared to a modern neighborhood built in 2010. Remember that all model estimates are subtracted from the different thresholds and that therefore a negative sign of B_1 is associated with an increase of the cumulative logit value. In summary, the negative coefficient implicates that if address density increases a neighborhood is less likely to have a high livability score and more likely to have a low livability score. This association is robust since I obtain negative coefficients for address density from an Ordinary Least Squares regression in table 7 in the appendix.

Moving onwards to the average marginal effects of address density I obtain positive coefficients for the first seven livability levels, but negative coefficients for the two highest livability levels as shown in figure 6 in the appendix. Indicating that when address density increases a neighborhood is more likely to end up with a low livability level than with a high livability level. This rejects the first hypothesis which states that a higher address density leads to higher livability levels. Moreover, table 8 in the appendix shows extreme small marginal effects for ending up with a livability level of 1 till 5. A reason might be that the histogram of the livability level is skewed to the left. This implies that there are few neighborhoods with low livability levels and many with high livability levels. The overall chance of ending up in a neighborhood with a low livability level is small, regardless of the number of addresses per square kilometer.

The second model in table 6 has multiple coefficients. Proximity of the railway station is now added into the regression because the literature tells that proximity public amenities might influence both address density and livability level. The coefficient B_1 has a significant value of -0.00045, indicating that the first estimated coefficient of address density contained an upward bias, but has the same magnitude. The average marginal effects show as expected that adding proximity of the closest railway station makes it more likely to find a neighborhood with a low livability score and less likely to find one with a high livability level when address density increases.

The third model represented in table 6 got proximity of the closest transfer station in the estimation. The previous estimated coefficient B_1 has a higher value for an increase in address density. The coefficient is equal to -0.00055. This proves the upward bias in line with the expectations in the section methods. Adding the variable proximity of a main road in the fourth model in table 6 results in an insignificant coefficient B_4. Furthermore, the coefficient of the main variable of interest address density in model 4 shows no change compared to model 3. Therefore, the third control variable has no value in the model. A reason might be the good national coverage of road infrastructure in the Netherlands which makes the distinctiveness of the variable low. The results with standard deviation and other statistics of model 1 till 4 are presented in table 9 till 12 in de appendix.

5.2 Poverty

Table 13: Ordinary Least Squares estimations with the determinants of households below social minimum in 2018.

	(1) Households Below Social Minimum	(2) Households Below Social	(3) Households Below Social	(4) Households Below Social	(5) Households Below Social Minimum	(6) Households Below Social
Address Density	0.00131*** (0.0000351)	0.00130*** (0.0000285)	0.000692*** (0.0000216)	0.000687*** (0.0000225)	0.000675*** (0.0000227)	0.000429*** (0.0000423)
Housing Price		-0.0139*** (0.000662)	-0.00156*** (0.000353)	-0.00158*** (0.000354)	-0.00147*** (0.000353)	0.0000929 (0.000449)
Public Rental Properties			0.141*** (0.00263)	0.141*** (0.00263)	0.141*** (0.00263)	0.148*** (0.00257)
Proximity Railway Station				-0.00405 (0.00433)	0.000631 (0.00455)	0.0160 (0.0120)
Dummy Railway Station					0.250*** (0.0899)	0.313*** (0.0930)
Constant	3.992*** (0.0528)	7.514*** (0.172)	1.986*** (0.117)	2.018*** (0.124)	1.928*** (0.126)	2.676*** (0.540)
R-Squared	0.2090	0.3371	0.6136	0.6141	0.6145	0.6976
Fixed Effects	NO	NO	NO	NO	NO	YES
N	9112	8961	8957	8953	8953	8953

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Regressing the independent variable address density on the dependent variable households below social minimum results in a positive first coefficient. The coefficient of 0.00131 tells that when the address density in a neighborhood increases with 1000, the percentage of households below social minimum increases with 1.3 percent point as displayed in table 13. The coefficient is significant on a 1% significance-level, whereby the confidence interval of address density [0.0012; 0.0014] is established with a method, which gives in 95% of the cases the unknown population average within the borders of the interval. Moreover, the interval shows that the coefficient is not equal to 0, which is also proved by the F-test. Therefore, the hypothesis that all coefficients are equal to zero is rejected.

The second regression model has a multiple format since average housing price is added. The coefficient of address density stayed the same looking at the four digits. This result shows not the expected upward bias of the previous estimated coefficient B_1. It might be that the average housing price does not influence the address density. However, the effect of housing price on households below social minimum is strong. The B_2 coefficient shows that when the average housing price increases with €100 000,-- the households below social minimum decrease with 1.4 percent point. This makes sense, because to buy a more expensive house, you need a higher mortgage and therefore a higher income. So, in a neighbourhood with more expensive houses, people are expected to have a higher income and thus there is a lower chance to find incomes below social minimum.

In the third regression model the percentage of public rental properties is added. The coefficient B_1 is now equal to 0.000692 on a 1 percent significance level. Indicating that a rise of 1000 addresses within a square kilometer leads to an increase of 0.6 percent point of households with an income below the social minimum. The expectation in the section methods that leaving out the variable public rental properties leads to an upward bias of the coefficient seems to hold. Meaning that the percentage of rental properties fades away a significant part of explanatory power of address density. Finally, the F-test tells once more that at least one of the coefficients is significantly different from 0.

The fourth regression model contains an additional control variable, the proximity of closest the railway station. The change of the B_1 coefficient is nihil and the coefficient B_4 is not significant. This is striking since the literature states that public transport amenities play a significant in declaring poverty in dense areas. Therefore, the fifth regression model contains a dummy coefficient B_5, which is equal to 1 when a station is reachable within 1.5 kilometer. Indeed, I see that the previous estimated coefficient B_1 in model 3 has a small upward bias, as expected. Furthermore, the dummy variable is now positive and significant on a 1% level. This implies that if a railway station lies within 1.5 km the number of households with incomes below the social minimum increases.

In model 6 in table 13 the municipality fixed effects are added into the regression. For the total of 380 municipalities, 379 municipality dummies are added which control for all time-irrelevant differences between municipalities. The coefficient of address density is now equal to 0.000429 on a 1% significance level. Implying that, when the address density increases with 1000 the households below social minimum increases with 0.4 percent point. The measured effect is small, possibly the variation within municipalities is small and most variation can be obtained between municipalities. However, this is now covered in the fixed effects. The F-test is not available for this regression, but the confidence interval shows that the coefficient is not equal to zero. The interval of address density [0.00035; 0.00051] is established with a method, which gives in 95% of the cases the unknown population average within the borders of the interval. Finally, the coefficient B_1 is positive in each of the six models, which is in accordance with the second hypothesis that a higher address density leads to more poverty.

5.3 Poverty and criminality conditional on address density

To test the third hypothesis the previous cross-sectional data in combination with the number of criminal incidents are used to test if more households living below the social minimum income level are associated with more criminal incidents. Furthermore, by adding address density and the interaction of households below the social minimum with density, I can measure if economies of scale in dense areas lead to a stronger association between poverty and criminality.

Table 14: Ordinary Least Squares regression estimations with interaction term.

	(1) Criminality in 2018	(2) Criminality in 2018	(3) Criminality in 2018
Households Below Social Minimum in 2018	1.049***	0.745***	1.149***
	(0.0606)	(0.0834)	(0.115)
Address Density		0.00191***	0.00370***
		(0.000278)	(0.000467)
Interaction Term			-0.000196***
			(0.0000349)
Constant	5.537***	4.483***	1.795***
	(0.331)	(0.312)	(0.649)
N	9089	9089	9089

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

In table 14 the three regressions have a positive coefficient for households below the social minimum income. This implies that a higher share of poor population is associated with more criminal incidents. The coefficients for households below the social minimum income level are for all regressions significant on a 1% significance level.

The coefficient in regression 1 is equal to 1.049. This implies that if the number of households below social minimum income increases by 1 percent point, there is at least one criminal incident per 1000 inhabitants more in the neighborhood. In the second regression address density has a positive coefficient indicating that when the address density is higher there are more criminal incidents per 1000 inhabitants. However, I obtain that the coefficient for poverty is smaller. This implies that this coefficient contained an upward bias in regression 1 but does not mean that more households below the social minimum lead to less criminal incidents since the coefficient B_1 is still positive. In the third regression in table 14 the interaction between poverty and density is added. The coefficient is significantly negative on a 1% level. This implies that when poverty and address density are both higher there are more criminal incidents, but there is no multiplier effect represented by the interaction term.

The third hypothesis, which states that the association between poverty and criminal incidents is stronger in dense neighborhoods is rejected. It is true that an increase in households below the social minimum income and that a higher address density are both associated with more criminal incidents. However, the interaction term is negative which implies that the association between poverty and criminal incidents is not stronger in dense neighborhoods. Instead, the association between households below the social minimum and criminal incidents itself gets stronger if we control for density.

6. Conclusion & Discussion

This paper investigates the relation of address density with livability and poverty. Firstly, the association between address density and livability is clarified by using several ordered logistic regressions. Secondly, the relation between address density and the percentage of households below the social minimum income is estimated. The statistical estimations are performed by using Ordinary Least Squares regressions with variables based on previous research of Glaeser et al. (2008). Finally, four quartile regressions show the association between poverty and the number of criminal incidents, an important determinant of livability, in neighborhoods conditional on address density. The data from Dutch neighborhoods in 2018 is used to answer the following research question:

What is the influence of address density on livability and how is poverty related to both in the Netherlands in 2018?

To answer the research question three hypotheses have been constructed. The hypothesis which states that a higher address density is associated with a higher livability level is rejected. The coefficient of address density in the ordered logistic regressions implies that an increase in address density is associated with a lower livability level. Two possible explanations might be that problems in dense areas in the Netherlands are not tackled enough or that side effects of a higher address density like more productive criminal activities decrease livability.

On the contrary, the second hypothesis is approved. I expected that a higher address density is associated with higher shares of poverty. The positive coefficient of address density implicates that when a neighborhood has more addresses per square kilometer, the percentage of households below the social minimum income increases. This confirms the theory of Glaeser et al. (2008) saying that poor people sort in places with good public transport amenities. Moreover, from my models it becomes clear that public housing has a larger predicting power than public amenities in the Netherlands.

Furthermore, the hypothesis which expects a stronger association between poverty and criminality in dense areas is rejected. Even if, poverty and density have both a positive coefficient when the interaction term is included in the regression. I do not obtain a stronger association in dense areas, referring to the negative interaction term. Despite that, the results approve with the idea that criminality arises more in poor and dense neighborhoods and is in line with the theory of Glaeser & Sacerdote (1999). They state that dense areas have scale economies which also hold for criminal activities.

In conclusion, a higher address density has a negative significant effect on the livability level, but a positive significant effect on the percentage of households below the social minimum income. The positive association of poverty with criminality in neighborhoods partially explains the lower livability in dense areas. This answer confirms the contradiction between on the one hand the demand to live in dense areas and on the other hand the lower livability and higher poverty levels. Therefore, the research is in contradiction with the theory of Von Thünen (1830), but in line with the research of Glaeser et al. (2008).

Even though, the performed Ordered Logistic regressions and Ordinary Least Squares regressions contain several control variables there might be problems with the internal validity. When these kinds of regressions have some Omitted Variable Bias, the results show a wrong relation between the independent and dependent variable. To solve this issue, more control variables can be added or a method which is less sensitive for Omitted Variable Bias can be used, such as an Instrumental Variable. Another potential problem is that some observations in the data contain missing values. The Percentage of households below the social minimum is not available for every neighborhood. It might be worthwhile to use incomes of individual people instead of households to do a similar analysis with more observations.

The external validity for a country like the Netherlands is high. The data is extracted from the Dutch statistical office and contains Dutch neighborhoods in 2018. However, for a foreign country the external validity is low. The treatment of poor people between countries is different and the Netherlands has relatively small cleavages between rich and poor in comparison with other countries. The fact that the data is reported recently is an advantage of this research.

This paper gives opportunities for future research. The results according to address density and livability might be estimated for different determinants of livability and between different provinces. People in dense areas have probably different preferences compared to people in rural areas. Furthermore, it is worthwhile to get the positive association between poverty and criminality causal by using time-series data to exclude reverse causality. Moreover, a lower geographical level can be estimated as a robustness check for the previous presented results, for example at street level.

The negative association between address density and livability is something to think about for municipalities which are planning to expand their cities to reduce the housing shortage. These municipalities must focus on methods to reduce the housing shortage while livability stays constant or improves. Besides that, the role of centralized poverty is important for rural, but even more for dense areas according to criminal incidents. Even if, the obtained relation is not causal yet, it supports the idea of the majors to invest in the most vulnerable neighborhoods of the Netherlands.

7. References

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

Balsas, C.J.L., (2010). Measuring the livability of an urban centre: an exploratory study of key performance indicators. *Planning, Practice & Research, 19*(1), 101-110. Doi: https://doi.org/10.1080/0269745042000246603

Bolt G., Burgers J. & Van Kempen, R. (1998). On the social significance of spatial location: spatial segregation and social inclusion. *Netherlands Journal of Housing and the Built Environment, 13*(1), 83-95. Consulted via https://link-springer-com.eur.idm.oclc.org/content/pdf/10.1007%2FBF02496935.pdf

Bolt, G. & Van Kempen, R. (2010). Escaping poverty neighborhoods in the Netherlands. *Housing, Theory and Society, 20(4)*, 209-222. Doi: https://doi-org.eur.idm.oclc.org/10.1080/14036090310018941

Camagni, R., Capello, R. & Nijkamp, P. (1997). Towards sustainable city policy: an economy–environment technology nexus. *Ecological Economics, 24*(1), 103-118. Doi: https://doi.org/10.1016/S0921-8009(97)00032-3

Centraal Bureau voor de Statistiek. (2019a). Kerncijfers wijken en buurten 2018. Geraadpleegd via https://opendata.cbs.nl/#/CBS/nl/dataset/84286NED/table?ts=1617805257342

Centraal Bureau voor de Statistiek. (2019b). Nabijheid voorzieningen; afstand locatie, wijk- en buurtcijfers 2018. Geraadpleegd via https://opendata.cbs.nl/#/CBS/nl/dataset/84463NED/table?ts=1617805944837

Cullen, J.B. & Levitt, S.D. (1999). CRIME, URBAN FLIGHT, AND THE CONSEQUENCES FOR CITIES. *The Review of Economics and Statistics, 81*(2), 159-169. Consulted via http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.381.9064&rep=rep1&type=pdf

Dinther, M. (2021). De grote steden gaan drastisch veranderen: minder auto's, meer hoogbouw en radicale vergroening. Consulted via https://www.volkskrant.nl/nieuws-achtergrond/de-grote-steden-gaan-drastisch-veranderen-minder-auto-s-meer-hoogbouw-en-radicale-vergroening~b45dab65/

Glaeser, E (2011). Triumph of the City: *How our Greatest Invention Makes us Richer, Smarter, Greener, Healthier and Happier.* Londen: Macmillan.

Glaeser, E.L., Gyouko, J. & Saks, R.E. (2006). Urban growth and housing supply. *Journal of Economic Geography, 6(1)*, 71-89. Consulted via https://www.jstor.org/stable/26160986

Glaeser, E.L., Kahn, M.E. & Rappaport, J. (2008). Why do poor live in cities? Consulted via https://www.nber.org/system/files/working_papers/w7636/w7636.pdf

Glaeser, E.L. & Sacerdote, B. (1999). *Why Is There More Crime in Cities?* Consulted via https://www.researchgate.net/profile/Bruce-Sacerdote/publication/24104112_Why_is_There_More_Crime_in_Cities/links/55b62f1a08ae9289a08aabb7/Why-is-There-More-Crime-in-Cities.pdf

Het Parool. (2008). Kwart alle armoede in grote steden. Consulted via https://www.parool.nl/nieuws/kwart-alle-armoede-in-grote-steden~ba3aa4f7/

Hooimeijer, P. & Van Kempen, R. (2000). De mogelijke doelen van beleid, In R. Van Kempen, *Segregatie en concentratie in Nederlandse steden: mogelijke effecten en mogelijk beleid.* Assen: Van Gorcum

Krugman, P. (2011). The new economic geography, now middle-aged. *Regional Studies 45*(1), 1–7. Doi: https://doi-org.eur.idm.oclc.org/10.1080/00343404.2011.537127

Leidelmeijer, K., Middeldorp, M. & Marlet, G. (2018). *Leefbaarheid in Nederland.* Consulted via Rapport (leefbaarometer.nl)

Margo, R. (1992) Explaining the Postwar Suburbanization of the Population in the United States; the Role of Income. *Journal of Urban Economics, 31*(3), 301-310. Doi: https://doi.org/10.1016/0094-1190(92)90058-S

Mieszkowski, P. & Mills, E. (1993). The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives, 7*(3), 135-147. Consulted via https://www.jstor.org/stable/2138447

Mills, E. & Lubuele, L.S. (1997). Inner Cities. *Journal of Economic Literature, 35*(2), 727-756. Consulted via https://www.jstor.org/stable/2729791

Muntendam, M. (2020). *De invloed van NS-treinstations op huizenprijzen.* Consulted via https://drive.google.com/file/d/1ETmI_jioYlei8fSARQkwRDdPrg7yMVhq/view?usp=sharing

Musterd, S. (2002). Response: mixed housing policy: a European (Dutch) perspective. *Housing Studies, 17*(1), 139-143. Doi: https://doi-org.eur.idm.oclc.org/10.1080/02673030120105956

Muth, R. (1968) Cities and Housing. Chicago: University of Chicago Press

Osofsky, G. (1966). *Harlem: The Making of a Ghetto.* New York: Harper & Row, Publishers.

Patterson, E.B. (1991). POVERTY, INCOME INEQUALITY, AND COMMUNITY CRIME RATES. *Criminology, 29*(4), 755–776. Doi: https://doi-org.eur.idm.oclc.org/10.1111/j.1745-9125.1991.tb01087.x

Poterba, J.M. (1984). Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach. *The Quarterly Journal of Economics, 99*(4), 729-752. Doi: https://doi-org.eur.idm.oclc.org/10.2307/1883123

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy, 82(1)*, 34-56. Consulted via https://www-journals-uchicago-edu.eur.idm.oclc.org/doi/abs/10.1086/260169?journalCode=jpe

Shafer, C.S., Koo Lee, B. & Turner, S. (2000). A tale of three greenway trails: user perceptions related to quality of life. *Landscape and Urban Planning, 49*(3-4), 163-178. Doi: https://doi.org/10.1016/S0169-2046(00)00057-8

The ministry of Interior Affairs and Kingdom relations. (2018) Leefbaarometer - online informatie over de leefbaarheid in alle buurten en wijken. Consulted via Leefbaarometer Home

Trouw. (2021). *Burgemeesters vragen een half miljard euro voor zestien kwetsbare gebieden.* Consulted via https://www.trouw.nl/binnenland/burgemeesters-vragen-een-half-miljard-euro-voor-zestien-kwetsbare-gebieden~bcc50446/?referrer=https%3A%2F%2Fwww.google.com%2F

Turok, I and McGranahan, G (2013) Urbanisation and economic growth: the arguments and evidence for Africa and Asia. *Environment and Urbanization 25*(2) 465–482. Doi: https://doi-org.eur.idm.oclc.org/10.1177/0956247813490908

Van Kamp, I., Leidelmeijer, K., Marsman, G. & De Hollander, A. (2003). Urban environmental quality and human well-being: Towards a conceptual framework and demarcation of concepts; a literature study. *Landscape and Urban Planning, 65*(1-2), 5-18. Doi: https://doi.org/10.1016/S0169-2046(02)00232-3

Van Lenthe, F.J., Brug, J. & Mackenbach, J.P. (2005). Neighborhood inequalities in physical inactivity: the role of neighborhood attractiveness, proximity to local facilities and safety in the Netherlands. *Social Science and Medicine, 60*(4), 763-775. Doi: https://doi.org/10.1016/j.socscimed.2004.06.013

Von Thünen, J.H. (1830). *Der Isolierte staat in Beziehung auf Landwirschaft und nationalokonomie.* München: Pflaum.

Wittebrood, K. (2000). Buurten en geweldscriminaliteit: een multilevel-analyse. Mens and Maatschappij, 75(2), 92-109. Consulted via https://rjh.ub.rug.nl/index.php/MenM/article/view/18599

8. Appendix

8.1 Figures

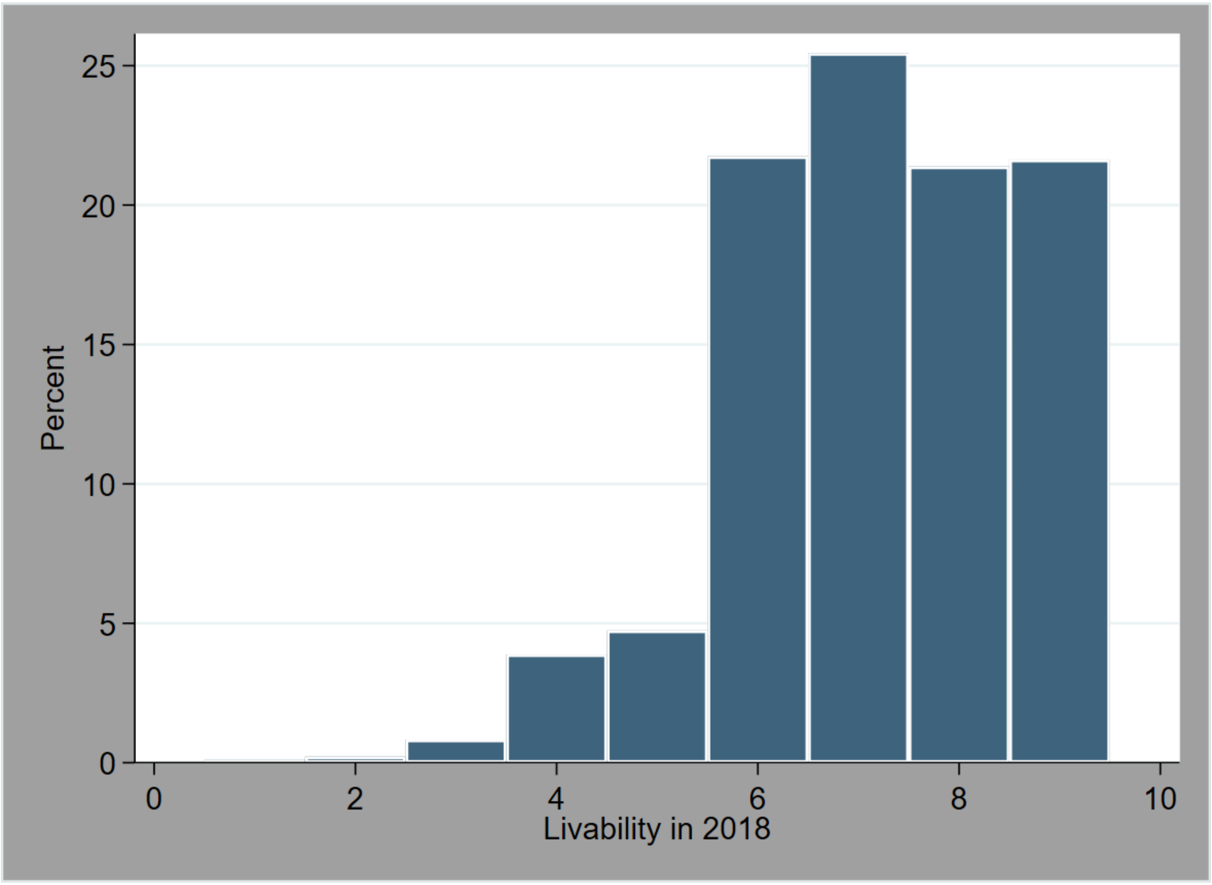

Figure 1: Histogram of livability in 2018.

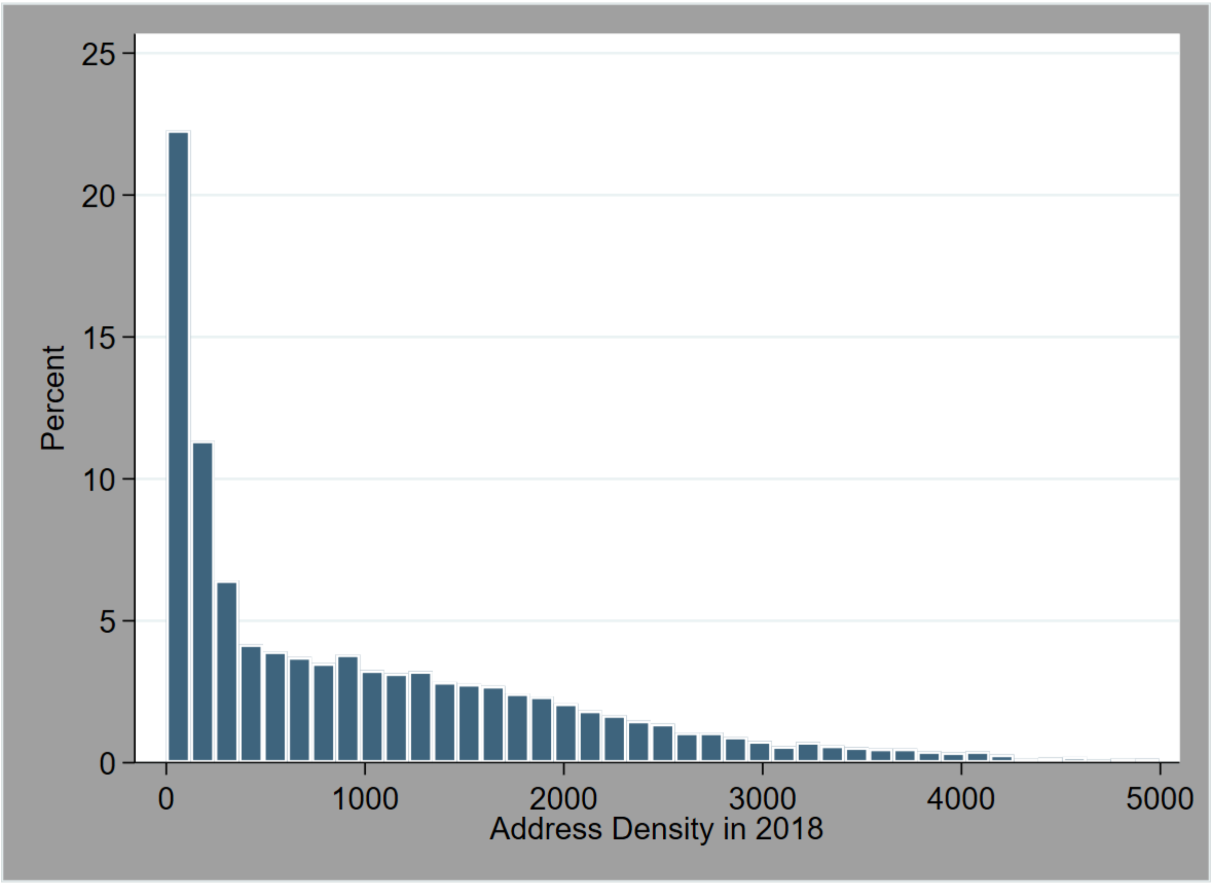

Figure 2: Histogram of address density in 2018.

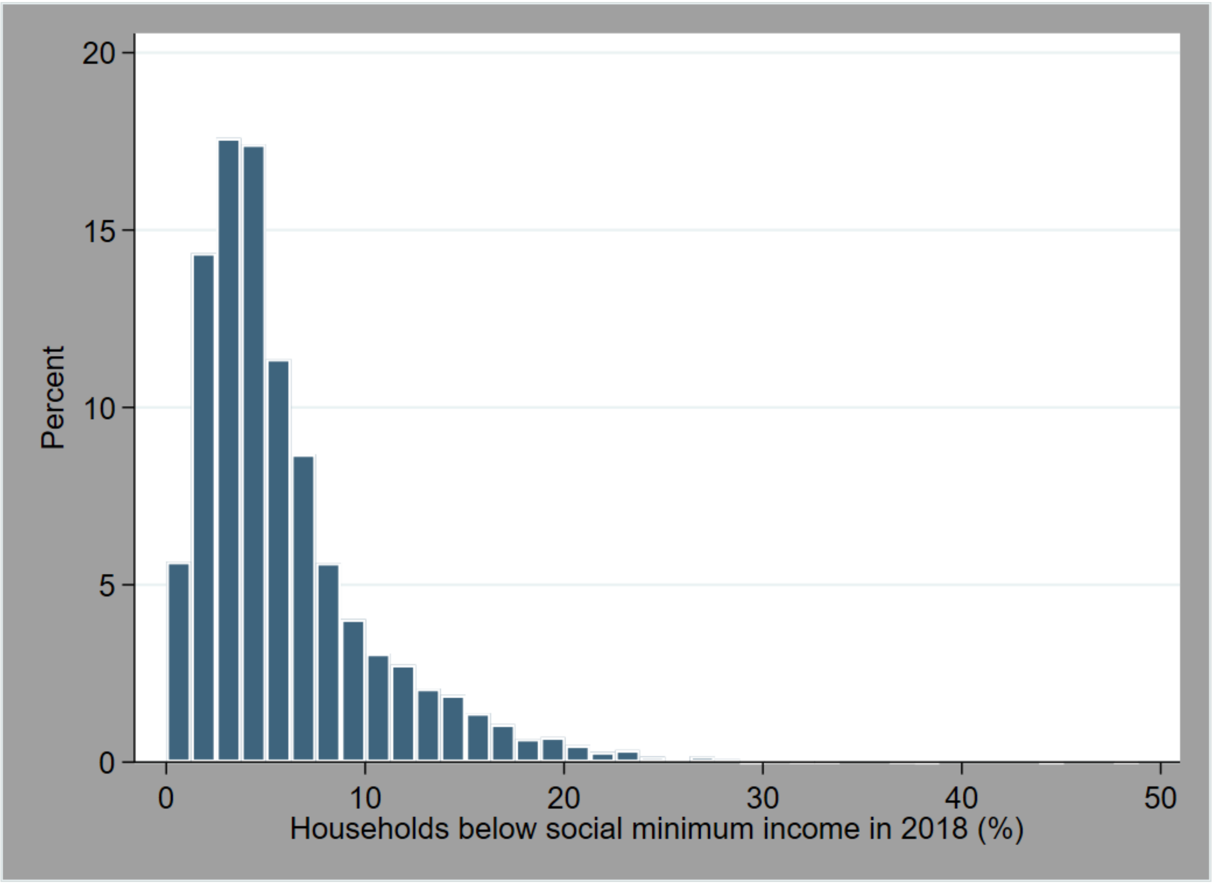

Figure 4: Histogram of Households below social minimum income in 2018.

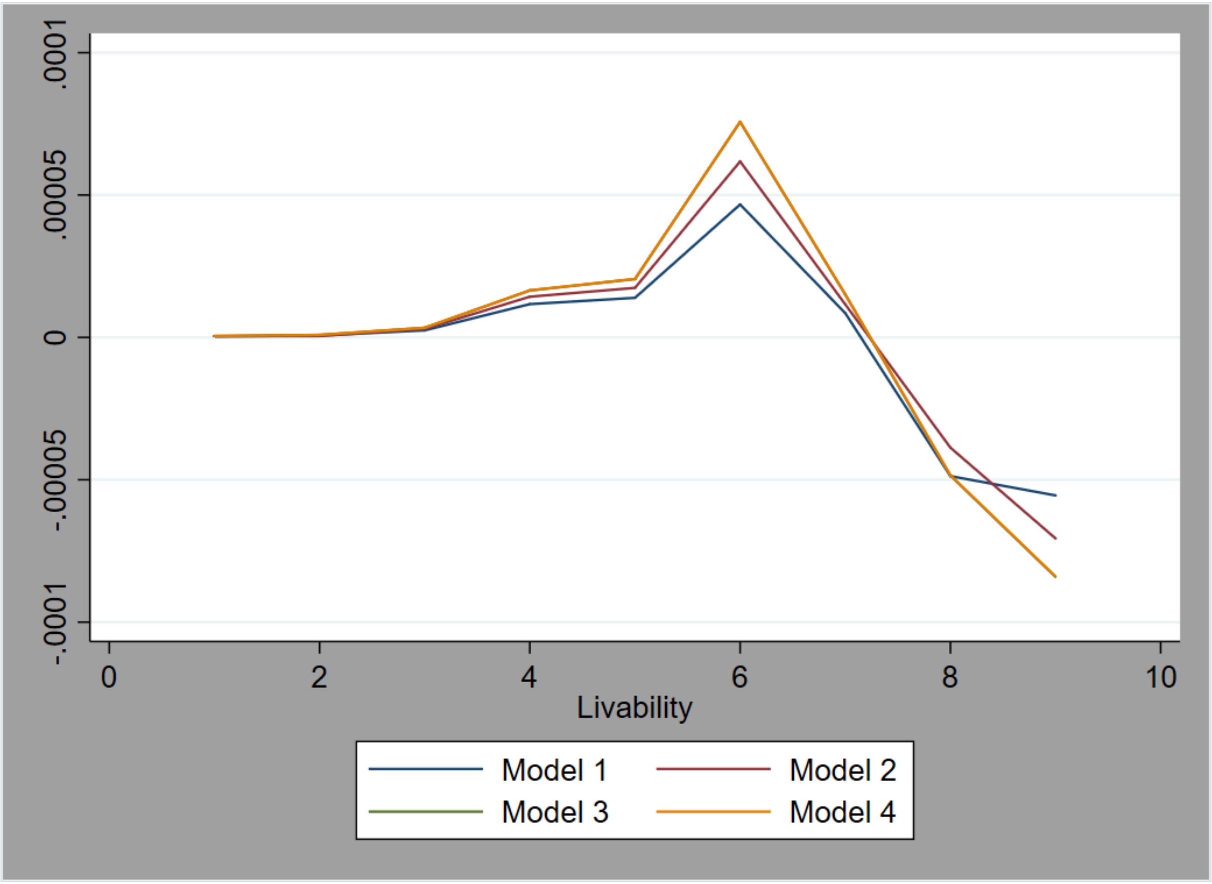

Figure 6: Average marginal effects of Address Density.

8.2 Tables

Table 3: Correlation coefficients performed for association with address density.

	Address Density	Livability level	Households below Social Minimum
Address Density	1		
Livability level	-0.24	1	
Households below Social Minimum	0.46	-0.6	1

Table 4: Name and unit of variables in the dataset.

Variable	Name	Year	Unit
kl_2018	Livability	2018	Ordinal Number
add_2018	Adress Density	2018	Square Km
prs_2018	Proximity Railway Station	2018	Km
ado_2018	Proximity Tranfer Station	2018	Km
pcr_2018	Proximity Main Road	2018	Km
hsm_2018	Households Below Social Minimum	2018	%
hsm_2017	Households Below Social Minimum	2017	%
hsm_2016	Households Below Social Minimum	2016	%
ahp_2018	Average Housing Price	2018	1000s Euros
rpv_2018	Public Rental Properties	2018	%
cri_2018	Criminal incidents	2018	Per 1000 inhabitants
cri_d_2018	Thefts	2018	Per 1000 inhabitants
cri_v_2018	Destruction	2018	Per 1000 inhabitants
cri_g_2018	Violence	2018	Per 1000 inhabitants

Table 5: Correlation coefficients performed for multicollinearity assumption.

	Adress Density	Railway Station	Transfer Station	Main Road	Housing Price	Public Rental Properties
Adress Density	1					
Railway Station	-0.33	1				
Transfer Station	-0.44	0.64	1			
Main Road	0.05	0.15	0.11	1		
Housing Price	-0.06	0.01	-0.06	0.04	1	
Public Rental Properties	0.038	-0.17	-0.20	-0.03	-0.48	1

Table 7: Ordinary Least Squares regression with the address density as determinant of livability level in 2018.

	(1) Livability	(2) Livability	(3) Livability	(4) Livability
Address Density	-0.000237***	-0.000286***	-0.000326***	-0.000325***
	(0.0000135)	(0.0000151)	(0.0000164)	(0.0000164)
Proximity Railway Station		-0.0336***	-0.0143***	-0.0141***
		(0.00171)	(0.00192)	(0.00193)
Proximity Transfer Station			-0.0242***	-0.0242***
			(0.00162)	(0.00162)
Proximity Main Road				-0.00466
				(0.00786)
Constant	7.477***	7.756***	7.998***	8.004***
	(0.0183)	(0.0266)	(0.0347)	(0.0366)
N	10986	10980	10980	10980

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Table 8: Marginal effects of address density in the ordinal logistic regressions.

Livability	dy/dx Adress Density	dy/dx Adress Density	dy/dx Adress Density	dy/dx Adress Density
1	0,000000352	0,000000421	0,000000475	0,000000475
2	0,000000651	0,000000478	0,00000088	0,000000879
3	0,00000248	0,00000298	0,00000338	0,00000337
4	0,0000117	0,0000143	0,0000165	0,0000165
5	0,0000139	0,0000174	0,0000205	0,0000205
6	0,0000467	0,0000618	0,0000757	0,0000757
7	0,00000853	0,0000116	0,0000151	0,0000151
8	-0,0000487	-0,0000387	-0,0000485	-0,0000484
9	-0,0000555	-0,0000706	-0,0000841	-0,000084
Controls		Railway Station	Railway Station Transfer Station	Railway Station Transfer Station Main Road

Table 9: Ordinal logistic regression with the address density as determinant of livability level in 2018.

Number of Obs	=	10986
LR chi2(1)	=	743.34
Prob > chi2	=	0.000
Pseudo R2	=	0.0202

Livability	Coef.	Std. Err.	z	P>|z|	95% Lower Bound	95% Upper Bound
Adress Density	-0,00035	0,00001	-25,99974	0,00000	-0,00037	-0,00032
Threshold 1	-7,35036	0,27896			-7,89711	-6,80361
Threshold 2	-6,30008	0,16707			-6,62754	-5,97262
Threshold 3	-5,04086	0,09307			-5,22327	-4,85845
Threshold 4	-3,49590	0,05011			-3,59412	-3,39768
Threshold 5	-2,74766	0,03926			-2,82460	-2,67072
Threshold 6	-1,21108	0,02689			-1,26377	-1,15838
Threshold 7	-0,09632	0,02404			-0,14343	-0,04922
Threshold 8	0,93454	0,02658			0,88246	0,98663

Table 10: Ordinal logistic regression with the determinants of livability level in 2018.

Number of Obs	=	10980
LR chi2(1)	=	1132.77
Prob > chi2	=	0.000
Pseudo R2	=	0.0308

Livability	Coef.	Std. Err.	z	P>|z|	95% Lower Bound	95% Upper Bound
Address Density	-0,00045	0,00002	-30,56465	0,00000	-0,00048	-0,00042
Railway Station	-0,05129	0,00256	-19,99485	0,00000	-0,05631	-0,04626
Threshold 1	-7,89992	0,28081			-8,45030	-7,34954
Threshold 2	-6,84920	0,17013			-7,18265	-6,51575
Threshold 3	-5,58791	0,09833			-5,78063	-5,39519
Threshold 4	-4,03243	0,05847			-4,14702	-3,91783
Threshold 5	-3,27469	0,04884			-3,37041	-3,17898
Threshold 6	-1,69618	0,03702			-1,76873	-1,62363
Threshold 7	-0,54802	0,03336			-0,61342	-0,48263
Threshold 8	0,50691	0,03429			0,43970	0,57413

Table 11: Ordinal logistic regression with the determinants of livability level in 2018.

Number of Obs	=	10980
LR chi2(1)	=	1455.69
Prob > chi2	=	0.000
Pseudo R2	=	0.0396

Livability	Coef.	Std. Err.	z	P>|z|	95% Lower Bound	95% Upper Bound
Adress Density	-0,00055	0,00002	-34,14609	0,00000	-0,00058	-0,00052
Railway Station	-0,02105	0,00305	-6,89846	0,00000	-0,02703	-0,01507
Transfer Station	-0,04138	0,00230	-17,96112	0,00000	-0,04589	-0,03686
Threshold 1	-8,43830	0,28292			-8,99280	-7,88379
Threshold 2	-7,38717	0,17356			-7,72734	-7,04700
Threshold 3	-6,12413	0,10402			-6,32799	-5,92026
Threshold 4	-4,55760	0,06664			-4,68821	-4,42699
Threshold 5	-3,79176	0,05782			-3,90509	-3,67842
Threshold 6	-2,18775	0,04671			-2,27929	-2,09620
Threshold 7	-1,01249	0,04255			-1,09588	-0,92909
Threshold 8	0,07000	0,04215			-0,01261	0,15261

Table 12: Ordinal logistic regression with the determinants of livability level in 2018.

Number of Obs	=	10980
LR chi2(1)	=	1455.83
Prob > chi2	=	0.000
Pseudo R2	=	0.0396

Livability	Coef.	Std. Err.	z	P>|z|	95% Lower Bound	95% Upper Bound
Address Density	-0,00055	0,00002	-34,05642	0,00000	-0,00058	-0,00052
Railway Station	-0,02091	0,00307	-6,80766	0,00000	-0,02693	-0,01489
Transfer Station	-0,04132	0,00231	-17,90065	0,00000	-0,04584	-0,03680
Main Road	-0,00387	0,01035	-0,37434	0,70815	-0,02415	0,01641
Threshold 1	-8,44439	0,28339			-8,99983	-7,88895
Threshold 2	-7,39324	0,17433			-7,73493	-7,05156
Threshold 3	-6,13015	0,10528			-6,33650	-5,92381
Threshold 4	-4,56313	0,06826			-4,69692	-4,42935
Threshold 5	-3,79695	0,05944			-3,91346	-3,68044
Threshold 6	-2,19272	0,04852			-2,28781	-2,09762
Threshold 7	-1,01751	0,04458			-1,10488	-0,93014
Threshold 8	0,06488	0,04428			-0,02190	0,15165

8.3 Coding

```
tabulate kl_2018 if sort=="Buurt     "
summarize add_2018 prs_2018 ado_2018 pcr_2018 hsm_2018 ahp_2018
rpv_2018 if sort=="Buurt     "
correlate add_2018 kl_2018 hsm_2018 if sort=="Buurt     "
histogram kl_2018 if sort=="Buurt     ", discrete percent
fcolor(edkblue) lcolor(white) lpattern(solid) ylabel(,
angle(horizontal)) graphregion(fcolor(gs10))
histogram hsm_2018 if sort=="Buurt     ", percent fcolor(edkblue)
lcolor(white) lpattern(solid) ylabel(, angle(horizontal))
graphregion(fcolor(gs10))
histogram add_2018 if sort=="Buurt     " & add_2018<5000, percent
fcolor(edkblue) lcolor(white) lpattern(solid) ylabel(,
angle(horizontal)) graphregion(fcolor(gs10))
twoway (scatter kl_2018 add_2018 if sort=="Buurt     " & add_2018<5000,
sort mcolor(edkblue) msize(small)) (lfit kl_2018 add_2018 if
sort=="Buurt     " & add_2018<5000), ylabel(, angle(horizontal))
graphregion(fcolor(gs10))
twoway (scatter hsm_2018 add_2018 if sort=="Buurt     " &
add_2018<5000, sort mcolor(edkblue) msize(small)) (lfit hsm_2018
add_2018 if sort=="Buurt     " & add_2018<5000), ylabel(,
angle(horizontal)) graphregion(fcolor(gs10))
summarize cri_2018 if sort=="Buurt     "


correlate add_2018 prs_2018 ado_2018 pcr_2018 ahp_2018 rpv_2018 if
sort=="Buurt     "
ologit kl_2018 add_2018 if sort=="Buurt     "
putexcel set "THS B5 Output Ologit 1", sheet("Blad1")
putexcel E1=("Number of Obs")     G1=(e(N))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A6=matrix(a), names
ologit kl_2018 add_2018 prs_2018 if sort=="Buurt     "
putexcel set "THS B5 Output Ologit 2", sheet("Blad1")
putexcel E1=("Number of Obs")     G1=(e(N))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A6=matrix(a), names
ologit kl_2018 add_2018 prs_2018 ado_2018 if sort=="Buurt     "
putexcel set "THS B5 Output Ologit 3", sheet("Blad1")
putexcel E1=("Number of Obs")     G1=(e(N))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A6=matrix(a), names
ologit kl_2018 add_2018 prs_2018 ado_2018 pcr_2018 if sort=="Buurt
"
putexcel set "THS B5 Output Ologit 4", sheet("Blad1")
putexcel E1=("Number of Obs")     G1=(e(N))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A6=matrix(a), names
eststo: regress kl_2018 add_2018 if sort=="Buurt     ", robust
eststo: regress kl_2018 add_2018 prs_2018 if sort=="Buurt     ", robust
eststo: regress kl_2018 add_2018 prs_2018 ado_2018 if sort=="Buurt
", robust
eststo: regress kl_2018 add_2018 prs_2018 ado_2018 pcr_2018 if
sort=="Buurt     ", robust
```

```
esttab using Output4.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear
ologit kl_2018 add_2018 if sort=="Buurt     "
mfx, predict(outcome(1))
mfx, predict(outcome(2))
mfx, predict(outcome(3))
mfx, predict(outcome(4))
mfx, predict(outcome(5))
mfx, predict(outcome(6))
mfx, predict(outcome(7))
mfx, predict(outcome(8))
mfx, predict(outcome(9))
ologit kl_2018 add_2018 prs_2018 if sort=="Buurt     "
mfx, predict(outcome(1))
mfx, predict(outcome(2))
mfx, predict(outcome(3))
mfx, predict(outcome(4))
mfx, predict(outcome(5))
mfx, predict(outcome(6))
mfx, predict(outcome(7))
mfx, predict(outcome(8))
mfx, predict(outcome(9))
ologit kl_2018 add_2018 prs_2018 ado_2018 if sort=="Buurt     "
mfx, predict(outcome(1))
mfx, predict(outcome(2))
mfx, predict(outcome(3))
mfx, predict(outcome(4))
mfx, predict(outcome(5))
mfx, predict(outcome(6))
mfx, predict(outcome(7))
mfx, predict(outcome(8))
mfx, predict(outcome(9))
ologit kl_2018 add_2018 prs_2018 ado_2018 pcr_2018 if sort=="Buurt
"
mfx, predict(outcome(1))
mfx, predict(outcome(2))
mfx, predict(outcome(3))
mfx, predict(outcome(4))
mfx, predict(outcome(5))
mfx, predict(outcome(6))
mfx, predict(outcome(7))
mfx, predict(outcome(8))
mfx, predict(outcome(9))
twoway (line Model1 Livability) (line Model2 Livability) (line Model3
Livability) (line Model4 Livability), graphregion(fcolor(gs10))

eststo: regress hsm_2018 add_2018 if sort=="Buurt     ", robust
eststo: regress hsm_2018 add_2018 ahp_2018 if sort=="Buurt     ",
robust
eststo: regress hsm_2018 add_2018 ahp_2018 rpv_2018 if sort=="Buurt
", robust
eststo: regress hsm_2018 add_2018 ahp_2018 rpv_2018 prs_2018 if
sort=="Buurt     ", robust
gen dprs_2018 = 1 if prs_2018<=1.5
replace dprs_2018 = 0 if prs_2018>1.5
eststo: regress hsm_2018 add_2018 ahp_2018 rpv_2018 prs_2018 dprs_2018
if sort=="Buurt     ", robust
```

```
eststo: regress hsm_2018 add_2018 ahp_2018 rpv_2018 prs_2018 dprs_2018
i.Municipality_id if sort=="Buurt     ", robust
esttab using Output1.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear

gen cri_2018 = cri_d_2018 + cri_v_2018 + cri_g_2018
gen hsm_add_2018 = hsm_2018 * add_2018
eststo: regress cri_2018 hsm_2018 if sort=="Buurt     ", robust
eststo: regress cri_2018 hsm_2018 add_2018 if sort=="Buurt     ",
robust
eststo: regress cri_2018 hsm_2018 add_2018 hsm_add_2018 if sort=="Buurt
", robust
esttab using Output3.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear
```