Understanding Housing Relocation Choice at the Neighbourhood-level

Laura Datta (Gärtner)
STUDENT NUMBER: u642187

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

dr. Katrijn van Deun
dr. Travis Wiltshire

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
17 July 2020

Acknowledgements

Writing this master thesis turned out to be very different than I had imagined. Instead of being at the office of the municipality of ’s-Hertogenbosch during the week, the Corona crisis and subsequent school closures forced my husband and me to work at home part-time, and cater to the needs of our two little girls 24/7. After some time of adjustment, we found a rhythm in splitting "working and caring" time equally among us, which required us to compensate for lost time in the evenings and weekends whenever possible. Nonetheless, I’ve made it!

I would not have been able to write this thesis without the extraordinary support of a few persons, which I would like to acknowledge next. Katrijn van Deun, my advisor at Tilburg University, was always very supportive of my situation and improved my thesis with excellent comments. Janine Meesters and Ronald Schouten from the municipality of ’s-Hertogenbosch provided valuable input to my thesis, even late in the evenings, struggling to find time themselves to get their work done. I would also like to express my gratitude to the O&S team at the municipality of ’s-Hertogenbosch for providing access to the data used in this thesis, and for the warm welcome in the team, even though we eventually did not see each other a lot.

I also wish to thank my parents for supporting my decision to return to University. I appreciate all your spontaneous travels all the way to the Netherlands to help us taking care of the kids! Lastly, a big thank you to you, Hannes. Without your encouragement to follow my ambition to become a data scientist, I would never have enrolled in this Master programme. Quite often, especially during exam periods, I could rely on you taking care of the girls, so I had enough time to study. And of course, thank you for being such an inspiring discussion partner during the thesis writing process. While you would often be very strict, it was because of you asking the right questions, giving the right comments and pushing me to the limits, that I’m eventually proud of this thesis, despite the circumstances I had to write it in!

Understanding Housing Relocation Choice at the Neighbourhood-level

Laura Datta (Gärtner)

Against the backdrop of housing shortages and rising housing prices, local governments need to understand how policies geared at growing the housing supply may affect consumer demand for new housing. Accordingly, this study proposes to analyse residential location choice at the neighbourhood level using detailed moving data for five years (2014 - 2018) for more than 15,000 households from the municipality of 's-Hertogenbosch. Based on a rich set of characteristics at the neighbourhood and household level, we first apply k-means clustering to identify neighbourhood types. Then, we estimate a conditional logit model which allows us to specify neighbourhood choice as a function of neighbourhood and household characteristics. In line with the results of previous studies, we find income to be an essential driver of neighbourhood choice. However, the role of ethnicity and household composition seems limited. Using our model results, we simulate the effects of several policy-relevant scenarios on the housing demand of households. We illustrate that, as the socio-economic situation of households deteriorates, the most vulnerable neighbourhoods risk further segregation in terms of income. We also show how potential policy measures may counteract this effect.

1. Introduction

1.1 Context

Housing shortage is a prevalent issue in the Netherlands. Market research estimated a shortfall of 3.8% of the total housing stock in 2019, implying a shortage of 294,000 houses (Kleinepier et al. 2019). The acute housing shortage is also considered one of the primary reasons for a surge in housing prices (de Groot and Vrieselaar 2019). Consequently, there is an urgent need for effective policies to stimulate growth in housing supply. Local governments play a crucial role in the creation of new living space. By selling or acquiring new ground, they can allocate land for housing construction. Through zoning plans, they can determine which type of dwellings to create for which target groups. Local decision-makers also determine the composition of the housing stock in terms of owner occupancy, private rent and social housing.

Housing policies can only be effective if they respond to the housing demand of consumers. Hence, local governments seek to be an attractive place for households already residing in the city as well as for households moving into the city from outside. Housing demand can manifest itself in two ways. First, by the need for a specific type of dwelling, e.g., a growing family with children requiring more space. Second, by the preferences for a particular residential location, e.g., that same family wanting to relocate to a neighbourhood with a high share of families with children.

Previous studies have highlighted the importance of neighbourhoods as the unit of analysis for residential location choices (Hedman, Van Ham, and Manley 2011; Clark, Deurloo, and Dieleman 2006; Van Ham, Boschman, and Vogel 2018; Mann et al. 2018). Neighbourhoods profoundly affect citizens in their everyday life: they determine perceived and actual safety of living surroundings, the availability of school and child care amenities, infrastructure quality, the appeal of exterior spaces, and the existence of a community feeling and mutual support among neighbours. Collectively, these factors make people feel 'at home', and hence directly affect their well-being (WRR 2005).

This study investigates residential location choices at the neighbourhood level. Households choose to relocate within the same or to different neighbourhoods based on their own characteristics (e.g., available income), and the characteristics of neighbourhoods (e.g., distance to train station). The interaction of household and neighbourhood characteristics is crucial to explain the neighbourhood choice process well (Hedman, Van Ham, and Manley 2011; Schelling 1971; Van Ham, Boschman, and Vogel 2018). For example, households with high income prefer to live in more affluent neighbourhoods (Hedman, Van Ham, and Manley 2011; Mann et al. 2018). Accordingly, the existing literature on neighbourhood choices has emphasised several drivers of neighbourhood mobility and neighbourhood selection at both the household and neighbourhood level (Hedman, Van Ham, and Manley 2011; Van Ham and Feijten 2008; Mann et al. 2018). Financial resources and the composition of the population in terms of ethnicity and household types are emphasised as the most significant factors influencing neighbourhood choice. As households appear to relocate to neighbourhoods with characteristics that match their own households characteristics, those studies suggest that neighbourhoods reproduce themselves over time. Local governments have a variety of tools available to influence housing demand at the neighbourhood level. Some of these tools - such as setting the share of social housing, or influencing the liveability by improving neighbourhood amenities - remain unaddressed in the literature. Therefore, in this study, we propose that the housing market structure (e.g., share of social housing, share of newly-built houses), the physical environment of the neighbourhood (e.g., distance to train station, number of restaurants nearby), as well as neighbourhood reputation are important drivers of neighbourhood choice. From the perspective of local policymakers, it is essential to understand what the implications of potential changes in those policies have on neighbourhood choices in specific, and residential mobility at large.

1.2 Research questions

This study seeks to assess why relocating households choose one particular neighbourhood type out of a set of alternative neighbourhood types.[1] To this extent, we zoom in on three particular research questions:

1. How can heterogeneous neighbourhoods be grouped together in a set of more homogeneous neighbourhood *types*?
2. *Which* factors explain households' choice for a particular neighbourhood type, based on the *interaction* of household and neighbourhood characteristics?

1 We acknowledge that an alternative research approach may seek to *predict* rather than *explain* neighbourhood choice. This is beyond the focus of this study. We discuss the merits of such a more exploratory approach in the discussion of our study.

3. *How much* do changes in household and neighbourhood characteristics influence households' choices for particular neighbourhood types?

In line with our research questions, we model neighbourhood choice in three steps. (1) First, as households choose neighbourhood *types* rather than specific neighbourhoods (Boschman and Van Ham 2015), we cluster comparable neighbourhoods based on a rich set of covariates using k-means clustering. The construction of neighbourhood clusters allows local governments to intervene with policies specific to a particular neighbourhood type across the entire municipality, rather than to focus on one single neighbourhood. (2) Next, we estimate a conditional logit model to explain differences in neighbourhood mobility between clusters. This type of discrete choice model allows us to analyse the specific neighbourhood choice of a household as a function of the characteristics of neighbourhoods. In order to understand variations between different types of households (e.g., single household or family with children), we interact the variables at the neighbourhood level with relevant explanatory variables at the household level, such as income, household type and age. (3) Using the results of the conditional logit model, we simulate how changes in either household or neighbourhood characteristics affect a household's neighbourhood choice process.

We calibrate our model on data from the municipality of 's-Hertogenbosch, The Netherlands[2]. Our principal data consists of household-level moving data based on the Dutch Personal Records Database (BRP) and the Key Registers Addresses and Buildings (BAG), which were provided by the municipality 's-Hertogenbosch for this study. We enrich our data set with neighbourhood characteristics from the Open Data Platform of the Dutch Statistics Bureau (CBS). Our workflow, encompassing data preparation, clustering, model estimation and simulation can be found at our public GitHub repository.[3]

1.3 Findings

After identifying nine neighbourhood clusters, we observe that (1) clusters with the lowest housing values and mean income are also the ones with the highest share of ethnic minorities and social housing. These clusters also receive the lowest scores on liveability. On the basis of a transition matrix, we illustrate that neighbourhoods have a high "staying power", i.e., many households moved within the same neighbourhood cluster and thus relocate to neighbourhoods with comparable characteristics to the one they are currently living in. (2) Based on the estimates of the conditional logit model, we find that household income is an important predictor of neighbourhood choice. Compared to findings reported in the literature, the role of ethnicity and household composition is less prominent. (3) In our simulation, we assess how a deterioration of the economic situation of households in the municipality affects vulnerable neighbourhoods, risking even further segregation. Households with little financial resources appear not to have the choice to relocate elsewhere but to remain in neighbourhoods with more affordable housing. We suggest two policies to counteract this effect (increase the availability of social housing, increase neighbourhood reputation), and measure the impact of these policies on the moving patterns of households.

2 's-Hertogenbosch is a mid-size city and municipality in The Netherlands, with approximately 150,000 inhabitants (2018).

3 See https://github.com/lauradatta/neighbourhood_choice. The data cannot be shared due to confidentiality agreements with the data provider.

The remainder of this thesis is structured as follows. The next chapter reviews the existing literature on residential mobility and neighbourhood choices. Subsequently, we introduce our modeling approach, which consists of k-means clustering and the conditional logit model. A description of the data and variables used for this study follows. In the subsequent chapters, we present the results of our modelling approach and illustrate implications by simulating various policies that could be enacted by local governments. We then discuss the findings as well as the limitations of our study, and suggest directions for future, more exploratory research. The last chapter concludes.

2. Literature review

Brown and Moore (1970) have identified two stages in a household's decision to move: the decision *whether* to move, and the decision *where* to move. Accordingly, there are two streams in the literature on residential mobility. The first stream focuses on explaining the decision process for households to seek a new residence (Coulter and Scott 2015). The second stream of literature zooms in on the household's decision of *where* to move (Hedman, Van Ham, and Manley 2011). This paper contributes to the second stream, by more closely investigating the decision of *where households move*. We review these literature streams next.

2.1 Why people move

Residential mobility has commonly been explained by the life course approach. The idea is that the decision to move is closely related to events in a person's life, such as marriage or divorce, the birth of children or the death of a family member (Coulter and Scott 2015; Dieleman 2001; Morris 2017; Thomas, Stillwell, and Gould 2016). Other important triggers are changes related to education and employment, which frequently require long-distance moves. The literature mostly considers age as a proxy variable for a household's mobility rate. Depending on different life stages, households have different propensities to move, e.g., young people entering adulthood might move more frequently to enter higher education or start their employment (Thomas, Stillwell, and Gould 2016). Next to age, household composition and income are considered as the most critical drivers of residential mobility. For example, a couple without children might be more flexible in moving than a family with children (Sánchez and Andrews 2011).

Residential relocation is also closely related to housing market conditions (Dieleman 2001; Sánchez and Andrews 2011; Van der Vlist et al. 2002). The national housing policies shape the housing market significantly. Through taxation policies, governments can influence the share of rental versus owner-occupied housing which can either increase or decrease residential mobility. Other instruments like rent control or tenant protection equally influence a household's decision to move (Sánchez and Andrews 2011). Finally, the structure of the local housing market influences mobility rates. Local governments set land use and allocation rules for new residential construction sites and determine the proportion of social housing in a neighbourhood (Van der Vlist et al. 2002), which lead to different mobility rates across local housing markets.

2.2 Where people move

After the initial decision to relocate, households typically choose the type and location of the new house, as well as the desired characteristics of the neighbourhood (e.g., proximity to public transport and services)[4] (Hedman, Van Ham, and Manley 2011).

2.2.1 Choice of housing type. Tenure (i.e., rent versus own), size and price form the essential characteristics for the investigation which types of dwellings households move to. One can observe a so-called 'housing career', i.e., an increase in housing size and property value as families grow and accumulate financial resources (Clark, Deurloo, and Dieleman 2006). Also, households consider whether properties are available for owning or renting in their decision to move. As home-ownership is a crucial source of a households' future financial assets, many households seek to become homeowner, especially in Europe and North America (Clark, Deurloo, and Dieleman 2006).

2.2.2 Drivers of neighbourhood choices. The process of neighbourhood choice is explained by a combination of household preferences (i.e., where households *want* to move to), and constraints (i.e., where households *have* to move to) (Brown and Moore 1970). The latter is determined by the availability of housing vacancies in a neighbourhood, the housing market structure, and the availability of financial resources. Jointly, those factors determine the feasible choice set for a particular household. In terms of setting preferences for specific neighbourhoods, households consider factors such as ethnic and socio-demographic composition, the physical environment as well as the reputation of the neighbourhood (Hedman, Van Ham, and Manley 2011; Mann et al. 2018; Van Ham, Boschman, and Vogel 2018). In the following, we review in more detail how those different drivers explain how households choose the neighbourhoods they move to.

Structure of the housing market. A key explanatory factor for neighbourhood choice is the availability of housing (Hedman, Van Ham, and Manley 2011). New dwellings become available either due to relocation or death or because new buildings were constructed. The process of relocation results in new vacancies, which form the basis of vacancy chains. Vacancy rates differ between neighbourhoods. For example, neighbourhoods with a high percentage of young and single adults have higher turnover rates than neighbourhoods with a high proportion of families with children (Van Ham and Clark 2009).

Financial resources. Financial resources available to a household are a crucial determinant for neighbourhood choices (Clark and Ledwith 2006). Accordingly, Hedman, Van Ham, and Manley (2011) find that a household's income explains much of the variation in neighbourhood choices. Low-income households do not relocate to neighbourhoods with a high median income, whereas high-income households are indeed more likely to move there. In this respect, the authors highlight that "choices are restricted by household preferences, resources, and restrictions, but also by the structure of the housing market" (Hedman, Van Ham, and Manley 2011, p. 1395).

4 While we acknowledge that the decision when *and* where to move can be made simultaneously, we assume a two-stage process for the remainder of this thesis.

Wealthier households usually have a more extensive set of neighbourhoods to choose from. Low-income households, instead, are confined to choose from a smaller range of neighbourhoods. As high-income households have more choices, they can more easily combine improvement in dwelling quality with better neighbourhood quality (Clark, Deurloo, and Dieleman 2006). By contrast, low-income households often only improve in terms of neighbourhood quality, and not in terms of housing quality.

Ethnic preferences. Several studies highlight the importance of the population composition and ethnic preferences in the choice for a neighbourhood (Boschman and Van Ham 2015; Hedman, Van Ham, and Manley 2011; Hedman and van Ham 2012; Van Ham and Clark 2009; Mann et al. 2018). Two processes can be observed: First, the white flight theory suggests that native people leave as the concentration of an ethnic minority in a neighbourhood increases. Similarly, the native population avoids those neighbourhoods in their choice process (Hedman and van Ham 2012; Van Ham and Clark 2009). Accordingly, Van Ham and Clark (2009) find that next to household income, the composition of the neighbourhood explains most of the differences between neighbourhood mobility in the Netherlands.

Second, households choose areas where the ethnic composition matches their own ethnic background (Schelling 1971). On the one hand, households express less willingness to leave their neighbourhood if the ethnic composition matches their own (Van Ham and Feijten 2008). On the other hand, households choose to relocate to neighbourhoods with an ethnic composition similar to their own ethnic background.

Socio-demographic factors. While income and ethnic composition are found to be the main drivers of neighbourhood choice, also socio-demographic factors such as level of education, household composition and age play a crucial role. For example, families with children may derive utility from living in neighbourhoods with a high percentage of parents (Hedman, Van Ham, and Manley 2011; Mann et al. 2018). In fact, several studies on neighbourhood choice suggest that neighbourhoods reproduce themselves over time - i.e., households relocate to neighbourhoods where the characteristics match their own household characteristics (Hedman, Van Ham, and Manley 2011; Ioannides, Zabel et al. 2008; Van Ham, Boschman, and Vogel 2018; Mann et al. 2018).

Neighbourhood location. The physical environment plays a crucial role in the location choices of households (Van Ham, Boschman, and Vogel 2018; Permentier, Van Ham, and Bolt 2009). Households decide to relocate not only because they want to improve in terms of the housing quality but also because they aim to live in a neighbourhood with improved quality (Clark, Deurloo, and Dieleman 2006). For example, access to public transport, such as proximity to the train station, and closeness to local amenities, such as shopping opportunities or restaurants, contribute to the attractiveness of a neighbourhood (Chhetri, Stimson, and Western 2006; Lee and Chun 2016).

Neighbourhood reputation. Studies suggest that also neighbourhood satisfaction and reputation drives moving decisions (Hedman and van Ham 2012). Accordingly, dissatisfaction with a neighbourhood increases the likelihood of moving away (Clark and Ledwith 2006). Permentier, Van Ham, and Bolt (2009) argue that not only neighbourhood satisfaction but also perceived reputation of a neighbourhood, i.e. how somebody believes others think about their neighbourhood, influences moving intentions.

2.3 Contribution to the literature

By investigating why households choose to relocate to a particular neighbourhood, we contribute to the current literature on residential location choices, and neighbourhood choices in particular. First, we combine comparable neighbourhoods into neighbourhood clusters, allowing local governments to address policies specific to a certain neighbourhood type across the entire municipality. In their decision on where to move, households usually choose a particular *type of neighbourhood* rather than a specific neighbourhood (Van Ham, Boschman, and Vogel 2018). While Clark, Deurloo, and Dieleman (2006) also identified neighbourhood types in their analysis, they do so relying on *predefined groups* based on socioeconomic and environmental status. We, instead, rely on an unsupervised machine learning algorithm, and include a comprehensive set of covariates that characterise neighbourhoods.

Second, existing literature on neighbourhood choice mainly has addressed how the interaction of financial resources and population composition in terms of ethnicity and household type at the neighbourhood and household level drive the neighbourhood choice process. From a policy-maker perspective, we believe that also other factors influence neighbourhood choice. We, therefore, extend previous discrete choice models on neighbourhood selection by also including variables describing the physical environment and the structure of the housing market. Through zoning plans and allocating space for social housing, local policy-makers can actively influence the composition of the housing market. Next to the more objective characteristics of neighbourhoods, we also include neighbourhood reputation as a subjective measure. Similarly to the housing market structure, local decision-makers can steer on the liveability of neighbourhoods not only through housing policies but also through other initiatives such as investing in the presence of police, or keeping the neighbourhood clean.

Third, local governments need to understand the potential effects of their policies on the housing demand of households. We broaden the existing literature on neighbourhood choices by simulating how changes in neighbourhood or household characteristics affect neighbourhood selection. We illustrate the direction and magnitude of the impact of such changes using several, policy-relevant scenarios.

3. Model

This study seeks to understand why households choose a particular type of neighbourhood. In line with our research questions, we apply k-means clustering to construct neighbourhood types. We then use those neighbourhood types to estimate a discrete choice model, based on characteristics at the neighbourhood and the household level.[5]

3.1 Clustering

We construct neighbourhood types using k-means clustering based on a rich set of neighbourhood characteristics (see Table 2). This unsupervised learning technique detects k distinct clusters of neighbourhoods by minimizing the total within-cluster

5 From a statistical point of view, it would have been possible to include *all* available neighbourhoods in the analysis, rather than only the more aggregated (clustered) neighbourhood *types* (for example, by using a random neighbourhood subset at the household level in the estimation procedure). However, our clustering approach eases the interpretation of our results for policy-making, and is therefore preferred.

variation, where each neighbourhood belongs to one cluster consisting of comparable neighbourhoods.

Clustering is an exploratory technique. As a result, the number of clusters is not known a priori. To determine the optimal value of k, i.e., the number of clusters, we rely on the Gap statistic. Specifically, Tibshirani, Walther, and Hastie (2001) propose to compare the total within-cluster variation for different values of k with their expected values under a null-reference distribution of the data, i.e., a distribution with no evident clustering. In line with the literature, we choose the value of k which maximises the Gap statistic.

3.2 Conditional logit model

Multinomial, discrete choice models explain the behaviour of individuals choosing from a set of given alternatives (in our case, neighbourhoods) (Mann et al. 2018). We now formalise this model, and summarise the model notation in Table 1. Following Greene (2012), for the ith household exposed to J choices, the utility U_{ij} of alternative j is:

$$U_{ij} = \theta Z_{ij} + \epsilon_{ij}$$

We assume that by choosing alternative j, individual i maximises U_{ij}, resulting in the following the statistical model:

$$Prob(U_{ij} > U_{jk}) \text{ for all } k \neq j$$

When specifying multinomial choice models, it is crucial to consider whether the utility of household i in choosing neighbourhood j depends on *household* characteristics (e.g., household income), the characteristics of the *neighbourhoods* (e.g., average dwelling values), or a combination of both (Greene 2012; Mann et al. 2018; Hoffman and Duncan 1988). It is therefore useful to partition Z_{ij} into household characteristics X_i and neighbourhood characteristics N_j, and split θ into $[\beta, \alpha]$. Traditional multinomial logit models specify the utility of choices by focusing on the characteristics of the individuals X_i. Variables which are specific to an individual are alternative-invariant; thus they do not vary across choices.

For modelling neighbourhood choice, however, the attributes of the alternatives, i.e., the characteristics of the neighbourhoods, are of key interest in the analysis. Explanatory variables describing the characteristics of the neighbourhoods N_j are varying over alternatives, i.e., they differ across neighbourhoods. The conditional logit model, as defined by McFadden (1974), allows us to model neighbourhood choice as a function of the characteristics of these alternative neighbourhoods. Assuming the independence of the error terms across alternatives J[6], the probability P_{ij} that household i chooses neighbourhood j, based on the attributes of the jth neighbourhood (N_j) and given the attributes of the other alternatives in the choice set N_k, is:

6 The idea of this assumption is that a household's unobserved preference for a certain neighbourhood is independent of its unobserved preference for an alternative neighbourhood. While we acknowledge the importance of this assumption, we do not address this issue in this study.

$$P_{ij} = \frac{\exp(\beta N_j)}{\sum_{k=1}^{K} \exp(\beta N_k)}. \tag{1}$$

Thus, the conditional logit model estimates the probability that, based on neighbourhood characteristics, a household chooses a specific neighbourhood (type) from a set of alternatives [7].

Table 1
Model Notation

Notation	Description
i	Household index (1, ..., I)
X_i	Characteristics of household i
j	Chosen neighbourhood
N_j	Attributes of chosen neighbourhood j
N_k	Alternative neighbourhoods

Next to the characteristics of the neighbourhoods N_j, neighbourhood choice also depends on the characteristics of the households X_i. In a location choice setting, the attributes of the alternatives N_j do only vary across alternatives, but not across households. This implies that for household $i = 1$, the characteristics of the neighbourhoods are the same as for household $i = 2, ..., I$ [8]. To include household characteristics in our model, we thus interact them with the neighbourhood characteristics. Formally, we extend equation 1 by interacting N_j with X_i to yield:

$$P_{ij} = \frac{\exp(\beta N_j X_i)}{\sum_{k=1}^{K} \exp(\beta N_k X_i)}. \tag{2}$$

To measure model fit, we use the McFadden pseudo-R^2, which compares the fitted model to a null model (Cameron and Trivedi 2005):

$$Pseudo - R^2 = 1 - \frac{logL_{fit}}{logL_0}. \tag{3}$$

4. Data

This study focuses on the municipality of ’s-Hertogenbosch, a mid-size city of about 150,000 inhabitants in The Netherlands. In 2018, ’s-Hertogenbosch counted 107 neighbourhoods.

7 Note, that the model is specified without an intercept as those cancel out.
8 Note that N_j only depends on j, and not on i.

4.1 Datasets

Data on neighbourhood characteristics. We obtain data on the characteristics of all neighbourhoods in ’s-Hertogenbosch from the Open Data Platform of the Dutch Central Statistics Bureau (CBS). Since the CBS does not publish data for neighbourhoods with less than approx. 30 households for data privacy reasons, we exclude 32 neighbourhoods where no data on household income and property value is available[9]. We further remove three neighbourhoods dominated by dwellings which do not fulfil a primary residential purpose (e.g., holiday parks). In total, we include 72 neighbourhoods in our analysis, which in 2018 account for a population of 70,300 households and 149,000 inhabitants.

Data on household relocations. The data on the moving behaviour of households for five years (2014-2018) is provided by the municipality of ’s-Hertogenbosch. We combine information from the Dutch Personal Records Database (BRP) and the Key Register Addresses and Buildings (BAG). The BRP contains personal data such as the birth of children, marriage and residential relocations on all individuals residing in the Netherlands. We aggregate the population data from the BRP at the level of the household head[10] and append it to information on dwellings with a residential purpose as registered in the BAG. The latter contains information on housing type, housing size and year of construction. By identifying address changes of household heads, we are able to track residential relocations of households.[11] For the purpose of this analysis, we focus on relocations within the municipality of ’s-Hertogenbosch, and from the region Noordoost-Brabant to the municipality.[12]

Our final dataset contains 34,031 moves. After deleting entries with missing data, we include 18,318 moves of 15,827 households in our analysis. On average, households move 1.16 times between 2014 and 2018.[13] Of those moves, 89% originate within the municipality of ’s-Hertogenbosch, and 11% occur from places *outside* of the region *into* the municipality.

Using our data set, we then derive a rich set of variables at the neighbourhood and household level to incorporate in our subsequent analysis. We give details on the operationalisation of all variables and their data sources in Table 2.

4.2 Clustering procedure

We apply k-means clustering on the covariates described in Table 3. We determine the optimal number of clusters at $k = 9$, by calculating the Gap statistic based on 50 bootstrapping samples, and a maximum number of 15 clusters. Subsequently, we cluster the data minimising the Euclidean distance within clusters, starting out with 25 initial

9 For about half of those 32 neighbourhoods, we also do not observe any residential moves in the data because these concern commercial and industrial areas.

10 For each household, CBS determines one household member as the household head.

11 We exclude moves resulting from death and any move where either household characteristics or data on the location of a household’s previous residence is missing.

12 Information on housing and neighbourhood choice is only provided to us for the region Noordoost-Brabant, requiring us to exclude moves from areas beyond these region from our analysis.

13 In our analysis, we treat moves as independent from preceding moves of the same household. First, households rarely move multiple times in our data. Second, household characteristics for each subsequent move *differ* from the characteristics of previous moves (e.g., in terms of the number of household members, or available space).

Table 2
Variable operationalisation

Dimension	Operationalisation	Source
(1) *Neighbourhood level*		
Financial resources	Mean neighbourhood household income in €, divided by 1000.	CBS[a]
Ethnicity	Share of native Dutch, non-Western and Western minorities as defined by CBS. Reference level: native Dutch.[e]	CBS[a]
Household type	Share of singles or others (single person households and multi-person households), couples (two person households), and families with children (including single parent households). Reference level: couples.[e]	CBS[a]
Housing market structure		
Average dwelling value	Housing value in €, divided by 1000, as determined by municipalities.	CBS[a]
Tenure	Share of home ownership, private rental (houses owned by private entities), and social housing (dwellings owned by social housing organisations). Reference level: share of home ownership.[e]	CBS[a]
Share of new houses	Share of dwellings built after the year 2000.[e]	CBS[a]
Physical environment		
Housing density	Average number of addresses per km^2 within 1-km radius.	CBS[a]
Distance to highway	Accessibility of neighbourhood, measured by the distance (in km) from the closest available highway access lane.	CBS[a]
Distance to train station	Centrality of neighbourhood location, measured as distance (in km) from the closest available central train station.	CBS[a]
Number of restaurants	Number of restaurants within a radius of 3 km.	CBS[a]
Neighbourhood reputation	Based on Dutch Liveability Index, measuring neighbourhood-level quality of life (housing, population, securities, amenities and physical environment), scaled between 1 (insufficient) and 9 (excellent). Updated every 2 years (2014, 2016, 2018). Values for 2015 and 2017 have been imputed by scores of preceding years.	BZK[b]
(2) *Household level*		
Type of move	Dichotomous variable, indicating whether move originated within the municipality, or outside of the municipality. Reference level: move within the municipality.	BRP[c]/ BAG[d]
Household income	Mean neighbourhood income by age groups, if not available filled with mean neighbourhood income. As 2018 values not available, imputed by using 2017 values.	CBS[a]
Ethnicity	Dichotomous variable, indicating whether household is native Dutch, Western or non-Western. Reference level: native Dutch.	BRP[c]
Household composition	Dichotomous variable, indicating whether household is single or other (single and multi-person households), couple (two persons household), or family with children (including single parent households). Reference level: couples.	BRP[c]
Age	Dichotomous variable, indicating whether household belongs to age category younger than 25, between 25 and 44, between 45 and 64 or older than 65 or not. Reference level: 25 - 44 years.	BRP[c]
Room stress	Number of household members, divided by m^2 of the dwelling preceding the move.	BRP[c]/ BAG[d]

Notes: [a] Data obtained through Open Data API Client of CBS, except for household income data (purchased by municipality ’s-Hertogenbosch from CBS.)
[b] Dutch Ministry of the Interior and Kingdom Relations.
[c] Dutch Personal Records Database, provided by the municipality ’s-Hertogenbosch.
[d] Key Register Addresses and Buildings, provided by the municipality ’s-Hertogenbosch.
[e] Shares are scaled between 0 and 100.

configurations. As common in k-means clustering, we standardise all neighbourhood

Table 3
Descriptive statistics: Neighbourhood characteristics (2018, N = 72 neighbourhoods)

	Mean	SD	Min	Max
Average dwelling values (x 1000, in €)	267	96.8	146	709
Mean household income (x 1000, in €)	30.6	6.9	18.8	49.7
Share of non-Western minorities (in %)	10.9	9.6	0	41
Share of Western minorities (in %)	8.5	2.6	1	13
Share of singles or other (in %)	36.8	16.2	12	80
Share of families with children (in %)	34.7	14.1	5	69
Share of private rental (in %)*	10.5	8.9	0	41
Share of social housing (in %)	34.6	26.2	0	95
Share of new houses	20.6	30.5	0	100
Housing density*	1,798.8	1,088.8	68	4,252
Distance to highway (in km)	1.9	0.7	0.6	4.5
Distance to train station (in km)	4.6	2.9	0.5	14.2
Restaurants within 3 km*	65.3	68.1	1	169.7
Neighbourhood reputation	6.8	1.5	4	9

* Excluded from the conditional logit model, as they were conceptually related to the variables *Share of social housing* and *Distance to train station*.

characteristics prior to clustering.[14] In the result section, we describe how we assign names to the neighbourhood clusters.

4.3 Conditional logit model

We standardise neighbourhood attributes and household characteristics[15], and then include their interactions in the conditional logit model (see Table 4 for descriptive statistics on the household characteristics).[16]

A residential move of a household to a new neighbourhood will affect the characteristics of the neighbourhood. As households decide to move based on the neighbourhood characteristics *before the move*, we always consider the neighbourhood characteristics at 1 January of the calendar year in which the move took place.

We use our model results to interpret the signs of the coefficients of the conditional logit model in section 5. To be able to not only assess the direction, but also quantify the magnitude of the effects, we *simulate* how changes in either the household or neighbourhood characteristics shift the probabilities of choosing neighbourhoods in section 5.3.

14 We initially applied our clustering technique separately for each year of our data. We found that neighbourhood clusters were very stable over time, meaning that neighbourhoods hardly switched between clusters from one year to another. We base our final neighbourhood clusters on the neighbourhood characteristics for the year 2018.

15 Standardisation was not used for dichotomous variables.

16 In order to reduce the complexity of the conditional logit model, we exclude three neighbourhood attributes from the model, that we *did* use in the clustering procedure: share of private rental, housing density, and the number of restaurants. Conceptually, these variables are represented by other variables (e.g., on the housing market and the location of the neighbourhood), not adding extra explanatory power to the model.

Table 4
Descriptive statistics: Household characteristics (Years 2014 - 2018, N = 18,318 households)

	Mean
Type of move	
Moves within municipality (in %)	88.5
Household income (x 1000, in €)	29.5
Ethnicity	
Native Dutch (in %)*	81.9
Non-Western minority (in %)	8.7
Western minority (in %)	9.5
Household composition	
Single or other (in %)	38.6
Couples (in %)*	31.3
Families with children (in %)	30.1
Age	
< 25 years (in %)	4.9
25 - 64 years (in %)*	55.2
45 - 64 years (in %)	24.1
> 65 years (in %)	15.8
Room stress	0.02

* Excluded from conditional logit model as reference levels.

5. Results

In this section, we first describe the results of the k-means clustering procedure. Next, we discuss the direction and significance of the coefficient estimates of the conditional logit model. Last, we describe our simulation approach and analyse the impact of changes in either household or neighbourhood characteristics on neighbourhood choices based on two example scenarios.

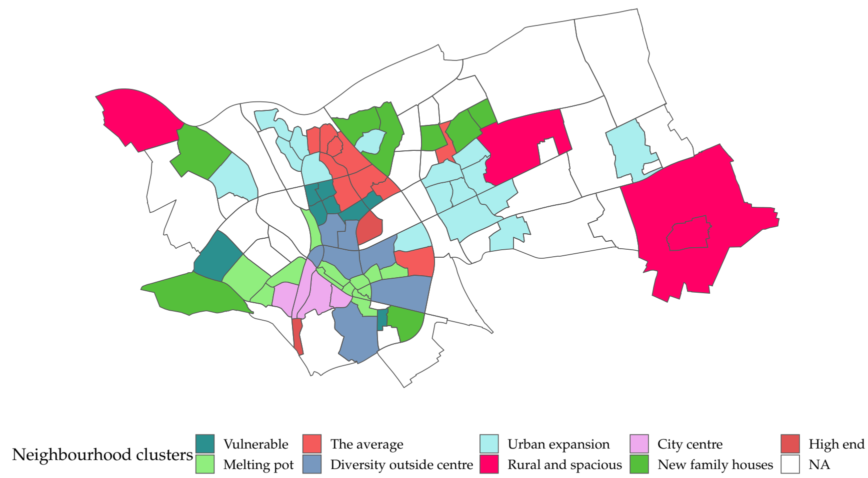
5.1 Neighbourhood clustering

We identify nine neighbourhood types through k-means clustering, as described in section 3.1. Figure 1 projects these clusters on a map of the municipality (colours represent the nine neighbourhood types). To ease the interpretation of our results, we assign names to each of the clusters, according to their most prominent characteristics. For example, the "melting pot" cluster receives its name from the diversity in terms of income levels and ethnicity. The "high-end" cluster, by contrast, is characterised by high income levels and housing prices[17].

5.1.1 Description of the neighbourhood clusters. Some neighbourhoods belonging to the same cluster are geographically located close to each other (e.g., "city centre", "diversity outside the city centre"). Other neighbourhood clusters spread across the municipality (e.g., "rural and spacious", "new family houses"). One important insight is that clusters mostly do not overlap with the administrative division of city districts, as defined by the municipality. This underscores the importance of relying on an unsu-

17 See Appendix A for a detailed description of the clusters. We have further verified and refined the naming of the clusters in discussions with policy makers at the municipality of ’s-Hertogenbosch.

Figure 1
Map of clustered neighbourhoods in ’s-Hertogenbosch



Note: Neighbourhood clusters identified through k-means clustering on the basis of neighbourhood characteristics measured in 2018. NA refers to neighbourhoods that were excluded from the analysis, either because of missing data, or because neighbourhoods did not have a primary residential purpose (e.g., holiday parks). Cluster names were assigned according to their most prominent neighbourhood characteristics (see footnote 17 for details).

pervised machine learning technique to group neighbourhoods in comparable clusters, rather than relying on a predefined classification.

Table 5 provides summary statistics for the neighbourhood types, based on a selection of characteristics[18]. We observe a positive relationship between average dwelling values and neighbourhood reputation, as it is the lowest for the "vulnerable" cluster (173,000€, with a reputation of 4.8) and the highest for the "high end" cluster (619,000€, with a reputation of 9). In addition, neighbourhood reputation seems to increase with the share of native Dutch households (58% for the "vulnerable" cluster, 94% for the "high-end" cluster). Ethnic minorities tend to live in clusters that are considered "less liveable", based on the subjective survey measure included in our study. In terms of household composition, the "new family houses" cluster is dominated by families with children (57%), while single households form the majority in the "city centre" neighbourhoods (63%). The "city centre" is located closest to the central train station (1.1 km), whereas the "rural and spacious" cluster is the furthest away (11.5 km).

Next, we provide insights on the moving patterns of households between the clusters.

Transition between neighbourhoods. Figure 2 shows a transition matrix between a household’s previous (i.e., before the move, shown in rows) and next (i.e., after the move, shown in columns) neighbourhood cluster. Using the figure, we learn which clusters households mainly move to, depending on their previous housing location. Specifically, the figure illustrates the *outward mobility* of neighbourhood clusters, expressed as the number of moves each new neighbourhood cluster "receives", divided by the sum of all

18 A table with *all* characteristics can be found in Appendix A.

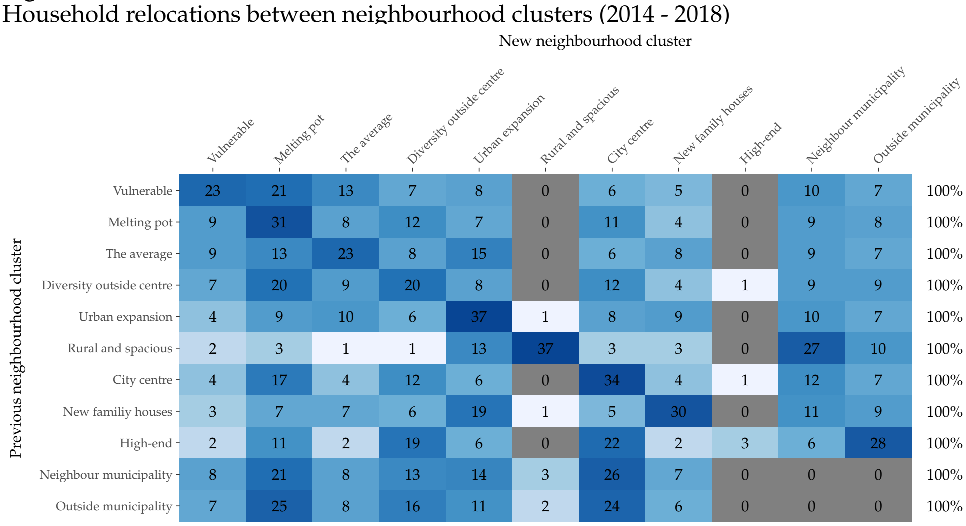
Table 5
Mean characteristics of neighbourhood clusters (2018, N = 72 neighbourhoods)*

Neighbourhood cluster	Reputation	Dwelling values (x 1000)	Native Dutch	Families with children	Social housing	Distance to train s tation
Vulnerable	4.8	173	58	39	75	4
Melting pot	5.3	189	74	21	63	2
The average	6.6	215	79	32	37	5
Diversity outside centre	6.6	250	83	26	29	2
Urban expansion	7.7	309	89	40	14	7
Rural and spacious	8	410	96	46	5	11
City centre	8	273	80	10	29	1
New family houses	8.6	321	86	57	18	6
High-end	9	619	94	44	0	2

* Table shows selected neighbourhood characteristics. A table with *all* variables, along with a detailed cluster description, is provided in Appendix A. Cluster names were assigned according to their most prominent neighbourhood characteristics (see footnote 17 for details).

moves from the originating cluster.[19]. Darker shading means a higher share of moves between clusters, while lighter shading means that fewer people relocate between clusters.

Figure 2
Household relocations between neighbourhood clusters (2014 - 2018)



Note: Values represent percentages of outbound moves, computed as the number of moves from a previous neighbourhood cluster to a new cluster, divided by the sum of all moves from each previous cluster. For example, 23% of all moves from the "vulnerable" cluster stay in the "vulnerable" cluster, while 21% of all moves from the "vulnerable" cluster relocate to the "melting pot" cluster.

19 Our model only includes moves *within the municipality* or originating *outside of the municipality*. For completeness, we also included moves ending *outside the municipality* here. We distinguish between neighbourhoods in municipalities bordering with ’s-Hertogenbosch, and the remaining "outside" municipalities, which are located further away.

For almost all clusters, the highest percentage of moves occur within the same neighbourhood cluster (on the diagonal in Figure 2). Households apparently stay in their neighbourhood, or move to comparable neighbourhoods. On the one hand, the "rural and spacious" cluster and the "urban expansion" cluster have the highest "staying power" (37%). On the other hand, the "high-end" cluster has the least staying power (3%). For the latter cluster, a majority of households does not stay within the municipality, but moves to a different one (28%). For the remaining clusters, the propensities to relocate outside of the municipality are much lower. The "rural and spacious" cluster forms an exception, as the majority of moves either occurs within the cluster, or to a municipality close-by (27%). The centrally-located clusters "city centre" and "the melting pot" are the most attractive neighbourhoods for new inhabitants of the city (i.e., households moving from outside of the municipality *into* the municipality), with only small differences between relocations from a neighbouring municipality (47%), compared to outside of the municipality (49%).

5.2 Results from the conditional logit model

So far, we have looked at the differences between clusters based on the characteristics of neighbourhoods. However, to accurately understand neighbourhood choices, we need to account for household-level differences in moving patterns. Therefore, we specify a neighbourhood choice model with interactions between households characteristics and the attributes of alternative neighbourhoods. We first discuss model selection, and then describe the parameter estimates of the conditional logit model.

5.2.1 Model selection. We initially specified a baseline model of neighbourhood choice in line with extant literature, interacting household income, ethnicity and household composition with their respective averages at the neighbourhood level (Hedman, Van Ham, and Manley 2011; Ioannides, Zabel et al. 2008; Mann et al. 2018). The results of this simple model largely confirm the findings reported in the literature. Households move to neighbourhoods matching their own characteristics, and thus, neighbourhoods tend to reproduce themselves over time.[20]

We extend the baseline model for two reasons: First, its pseudo-R^2 was quite low (3%, compared to a null model which only includes the interaction between neighbourhood and household income as an explanatory variable). Second, as explained in the introduction, the baseline model lacks a set of important covariates that describe households and neighbourhoods, and are essential from the perspective of local governments to understand neighbourhood choice. Therefore, we add to the baseline model variables on the location of the neighbourhood, the structure of the housing market and a measure for a neighbourhood's reputation. The explanatory power of the model increases to a pseudo-R^2 of 12.5%, justifying the inclusion of these extra covariates. We subsequently report the findings of this model in the remainder of this thesis.

5.2.2 Parameter estimates of the main model. In what follows, we report the results of the conditional logit model. We focus on a set of common themes that emerge from

20 For example, households with a non-Western migration background are more likely to relocate to neighbourhoods with a high share of non-Western minorities, compared to native Dutch households ($\beta = 0.306, p < 0.01$). Families with children are more probable to move to places where many families with children lived, compared to couples ($\beta = 0.221, p < 0.01$). See Appendix B for the full results of this model.

the results, rather than discussing each parameter estimate separately. Table 6 provides a summary of the results *in matrix form*, whereby coefficients represent the estimated interaction effects and significance levels between household characteristics (in rows) and neighbourhood characteristics (in columns). We provide a full set of results with standard errors in Appendix B.

Comparison between outside and inside moves. We find a distinct pattern between moves that originate *within* the municipality, compared to moves originating *outside* the municipality (i.e., households moving to the city from elsewhere). Households that stay in the municipality tend to move away from the city centre ($\beta = 0.214, p < 0.01$) to neighbourhoods with a higher share of social housing ($\beta = 0.292, p < 0.01$), or a higher share of newly built dwellings ($\beta = 0.327, p < 0.01$). Households that relocate to the municipality from elsewhere, instead, move to neighbourhoods with a high share of Western migrants ($\beta = -1.044, p < 0.01$), singles ($\beta = -0.345, p < 0.01$) and families with children ($\beta = -1.150, p < 0.01$). While moves into the city are more likely to happen to neighbourhoods with high dwelling values ($\beta = -2.616, p < 0.01$), they are less probable to occur to neighbourhoods with high neighbourhood income ($\beta = 2.027, p < 0.01$).

Those results underscore the observed pattern in the transition matrix provided in Figure 2. The most centrally located clusters are the most attractive ones for new inhabitants of the municipality. Moves *within* the municipality occur more towards neighbourhoods at the periphery.

Household income. In line with previous studies, the outcomes of the conditional logit model confirm the importance of household income in explaining neighbour choice (Clark and Ledwith 2006; Hedman, Van Ham, and Manley 2011; Van Ham, Boschman, and Vogel 2018). Most of the interactions with neighbourhood characteristics are significant and directionally face-valid. For example, richer households move to neighbourhoods with higher mean average incomes ($\beta = 0.342, p < 0.05$) and higher reputation ($\beta = 0.334, p < 0.01$). Richer households also have a higher probability of moving to the outskirts of the city, where houses are usually bigger (distance to highway: $\beta = 0.085, p < 0.01$). We find that with increasing household income, households are more likely to move to neighbourhoods with a lower share of Western minorities ($\beta = -0.224, p < 0.01$).

One particular finding warrants further explanation. Households with higher mean income seem to prefer neighbourhoods with *lower* average dwelling values ($\beta = -0.399, p < 0.01$). This seems counter-intuitive, as households should be able to afford more expensive housing with an increase in available income. However, we need to acknowledge that the household income variable is often approximated by the mean neighbourhood income in a particular age group, which may be insufficient to capture the actual income distribution of households in a particular neighbourhood.

Other variables. We observe that families with children selected locations outside the city centre, which are closer to the highway ($\beta = -0.401, p < 0.01$), further away from the central train station ($\beta = 0.397, p < 0.01$) and have a high share of newly-built dwellings ($\beta = 0.197, p < 0.01$). This finding suggests that, in general, families are attracted to neighbourhoods at the periphery, where new and more spacious family houses have been built.

Contrary to other studies on neighbourhood choice, non-Western minorities do not exhibit a significantly different choice pattern than native Dutch, except for a higher probability to move to neighbourhoods with social housing ($\beta = 0.712, p < 0.01$)

(Boschman and Van Ham 2015; Hedman, Van Ham, and Manley 2011; Van Ham and Feijten 2008). Western minorities are significantly more attracted by neighbourhoods which are close to the city centre ($\beta = -0.224, p < 0.05$) and move to neighbourhoods with low average household income ($\beta = -1.185, p < 0.05$), as well as a high share of non-Western minorities ($\beta = -0.222, p < 0.05$). In terms of household age, households older than 65 years have a higher probability of relocating further away from the city centre (i.e., further away from the train station), compared to mid-age households ($\beta = 0.280, p < 0.01$).

We observe that households with higher room stress, i.e., households having relatively fewer square metres per household member available, have a higher probability to relocate to places with a high share of singles ($\beta = 0.064, p < 0.1$). As those neighbourhoods usually are dominated by small houses, households with more need for space tend to move away. We also found that with increased room stress, households are moving to neighbourhoods where more social housing is available ($\beta = 0.162, p < 0.05$), which commonly are the most vulnerable clusters. A possible explanation is that, on average, with increased room stress, households cannot afford to move to neighbourhoods with bigger houses, because average dwelling values are too high.[21]

Above outcomes suggest that there is no single, distinct pattern of neighbourhood choice for the municipality of ’s-Hertogenbosch. While many of the interaction terms with the household characteristics "moves within municipality" and "household income" were face-valid, the coefficient estimates for the remaining variables showed a nuanced pattern, which calls for a careful analysis of the joint impact of changes to these variables on neighbourhood choice. Therefore, in the next section, we use our model to simulate how changes in the characteristics at the household and neighbourhood level affect the direction and magnitude of neighbourhood choice probabilities of households.

21 We simulate the effect of an increase in room stress on neighbourhood choice, and find a higher probability to move to the "vulnerable" and "melting pot" clusters for the average household. If, however, we *also increase household income* in the simulation, the average household becomes more likely to move to neighbourhoods with higher dwelling values, such as the "new family houses" cluster. The results of this simulation are provided in Figure 7 in Appendix C.

Table 6
Results of the conditional logit model (dependent variable: Neighbourhood choice)

Household characteristics	Dwelling value	Neighbourhood income	Non-Western minorities	Western minorities	Singles or other	Families with children	Social housing	New houses	Distance to highway	Distance to train station	Neighbourhood reputation
Moves within municipality [a]	-2.616 ***	2.027 ***	0.062	-1.044 ***	-0.345 ***	-1.150 ***	0.292 ***	0.327 ***	-0.437 ***	0.214 ***	0.120
Household Income	-0.399 ***	0.342 **	0.041	-0.224 ***	-0.063	-0.223 ***	-0.014	0.025	-0.085 ***	0.005	0.334 ***
Non-Western minority [b]	-0.484	0.541	0.073	-0.038	0.073	-0.020	0.712 ***	0.135	-0.051	-0.003	0.120
Western minority [b]	0.280	-1.185 **	0.222 **	-0.181	-0.771 **	-1.035 **	-0.420 **	0.292 **	-0.168 *	-0.224 **	0.281
Singles or other [c]	-1.108 ***	0.809 **	-0.057	0.021	-0.178	-0.319 *	0.183	0.009	0.171 ***	0.080	-0.240
Families with children [c]	-0.386 *	0.608 **	-0.071	-0.270 **	0.350 ***	0.119	-0.040	0.197 ***	-0.401 ***	0.397 ***	-0.274
< 25 years [d]	-0.412	-0.765	0.019	-0.009	0.560 *	0.491	-0.031	-0.421 ***	0.194 *	-0.217	0.980 ***
45 - 64 years [d]	-0.195	-0.437	-0.067	0.092	0.296 **	0.583 ***	-0.167	-0.241 ***	0.137 ***	-0.196 ***	0.049
> 65 years [d]	-1.462 ***	0.456	-0.134 *	-0.350 **	-0.313	-0.390	0.214	-0.236 **	-0.130 *	0.280 ***	0.328
Room-stress	-0.073	-0.009	-0.064 *	0.048	-0.152 *	0.002	0.162 **	-0.103 **	0.079 **	-0.104 ***	0.028

Notes: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

[a] Reference level: Moves into the municipality.
[b] Reference level: Native Dutch.
[c] Reference level: Couples.
[d] Reference level: 25 - 44 years.

The dependent variable measures a household's choice for one of the nine neighbourhood types included in the model. The table reports the interaction effects between household characteristics (in rows) and neighbourhood characteristics (in columns). Full model results, including estimated standard errors, are available in Appendix B.

5.3 Simulation: The effect of housing policies

Local decision-makers are interested in how changes in the socio-economic circumstances of households and specific housing policies affect moving behaviour in their municipality. Our model allows us to simulate these effects. Next, we discuss two example scenarios to illustrate the potential impact, both in terms of direction and size, on neighbourhood choices.

5.3.1 Simulation procedure. We first construct an "average household" (income of 29.500€, single, native Dutch, 25 - 44 years old, room-stress of 0.021)[22]. Based on the coefficients of the conditional logit model, we then predict the probabilities of relocation to any of the nine neighbourhood clusters.

For each of the subsequent simulation scenarios, we alter either the household characteristics of the relocating household, or the neighbourhood attributes of the potential target neighbourhoods. We again predict the household's choice probabilities for all nine neighbourhood clusters, and summarise our results in bar plots. Throughout, these plots show *absolute choice probabilities* for the different scenarios. The change, expressed in percentages (not percentage points) of pre- and post-choice probabilities is also shown in these plots, and is our key measure to assess the impact of a particular scenario on neighbourhood choice.

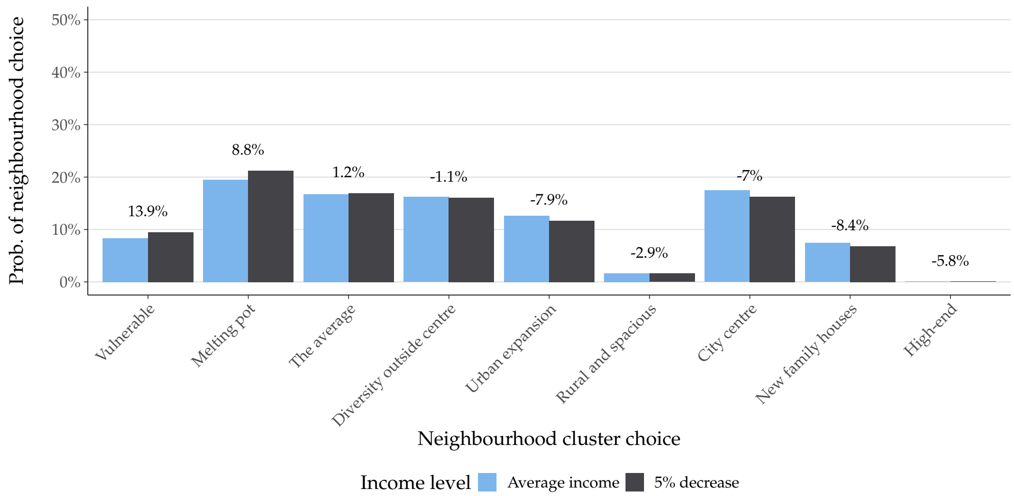
5.3.2 Simulation 1: Change in socio-economic situation of the household. From our results, household income has emerged as a central variable in predicting neighbourhood choice. However, due to the various interaction effects between income and neighbourhood characteristics, it is difficult to "spot" how a change in income would affect overall choice probabilities.

Simulation 1a: Response to negative income shocks. One interesting scenario to consider is how households' relocation choices are affected in times of economic downturns. To shed light on this issue, we simulated how an "average household" responds to a 5%-decrease in income. Does such a decrease affect moving patterns in a sizeable way? And if it does, which neighbourhood clusters (i.e., parts of the municipality) are affected most?

Figure 3 plots the *choice probabilities* for the nine neighbourhood clusters, for an average household (blue), and for the same household with a 5%-decrease in income (grey). From the chart, one can easily spot that choice probabilities for the "vulnerable" and "melting pot" clusters increase. These are both clusters with below-average dwelling values, which become more affordable during economic downturns. For example, the "vulnerable" cluster experiences a choice increase of 13.9% (from 8.3% to 9.5%, approx. 44 additional moves[23]), compared to a scenario where household income stayed the same. Neighbourhoods with higher housing prices, in turn, experience a decrease of moving probabilities. From Figure 3, we learn that especially the "urban expansion", "city center", and "new family houses" clusters may suffer from such negative income shocks.

22 Concerning the categorical variables household type, ethnicity and age category, we chose the category with the highest proportion in our data.
23 316 moves (yearly average in the "vulnerable" cluster) x 13.9% = 44 additional moves.

Figure 3
Simulation 1a: Response to a 5%-decrease in household income on neighbourhood choice



Note: The figure illustrates the probabilities of moving to one of the nine neighbourhood clusters for an average household (blue bars), and the same average household with a 5% lower income (grey bars). Labels in the plot show the percentage change in estimated probabilities between the blue and grey bars.

Simulation 1b: Testing the "white flight" theory. The current simulation setup also allows for a test of the so-called "white flight theory" in the municipality of 's-Hertogenbosch. Recall that this theory stipulates that native Dutch households may leave neighbourhoods with a high concentration of ethnic minorities if the opportunity to do so arises. Specifically, we assume that household income increases by 5%, and hence offers households the "opportunity" to move elsewhere.

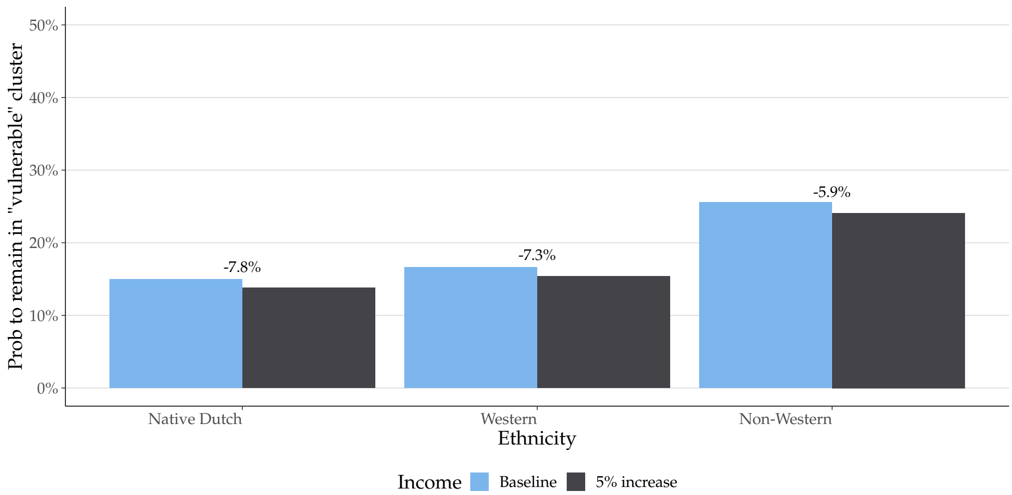
We test the white flight theory by simulating the relocation choices of residents of the "vulnerable" cluster, consisting of the most ethnically-diverse neighbourhoods (income of 22,100€, single, 25-44 years old, and room-stress of 0.019). We contrast the choice probabilities of native Dutch households, with those of two ethnic minorities. If the white flight theory holds, we would expect native Dutch households to experience a faster *decrease* in choice probabilities for their current neighbourhood, or - in other words - less "staying power", compared to their otherwise equal, but ethnically different, counterparts.

Figure 4 shows the difference in moving probabilities between native Dutch, Western and non-Western households. We notice that the impact of a growth in income is very similar for all three groups. The probability of relocating to the "vulnerable" cluster, which consists of neighbourhoods with the highest share of ethnic minorities, indeed reduces the least for households with a non-Western migration background (5.9%). While this finding is in line with the white flight theory, differences between ethnic groups remain relatively small in magnitude.

5.4 Simulation 2: Policies to diversify neighbourhoods

From above simulations, we find that vulnerable neighbourhoods (e.g., those with low neighbourhood reputation and high concentration of ethnic minorities) are most likely to suffer from the deterioration of economic circumstances, potentially due to low

Figure 4
Simulation 1b: Response to a 5%-income increase on probabilities to stay in the "vulnerable" cluster, based on ethnicity



Note: The figure illustrates the probabilities of an average household from the "vulnerable" cluster to remain in the same cluster (blue bars), and for the same household with a 5%-increase in income (grey bars), for three ethnic groups: native Dutch, Western, and non-Western. Labels in the plot show the percentage change in estimated probabilities between the blue and grey bars.

housing prices and the widespread availability of social housing. As a consequence, we might observe an increasing number of households with low incomes move to these neighbourhoods, leading to even more segregation *across* neighbourhoods in the long term. Local governments probably want to address this issue through their housing policies, which we evaluate next.

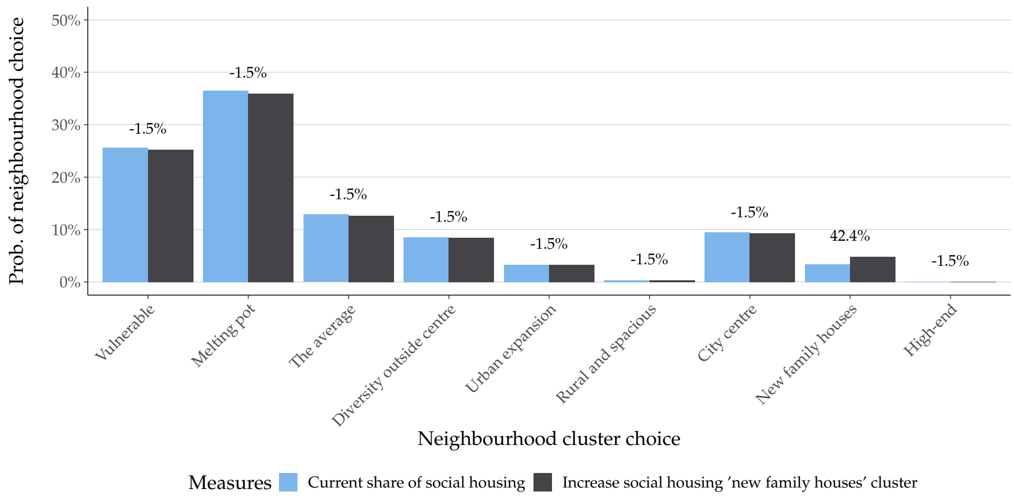
Simulation 2a: Decreasing neighbourhood stickiness. One possible solution to neighbourhood segregation is to reduce the "stickiness" of vulnerable neighbourhoods. Recall that a large share of households tends to *remain* in their originating cluster when moving, which keeps these neighbourhood from diversifying over time.

The conditional logit model allows us to estimate the probability of a household to remain in the same cluster.[24]. We simulate how households from the "vulnerable" cluster (for average household characteristics see section 5.3.2) respond to an increase in the share of social housing in the "new family houses" cluster (we set the share of social housing to 25% - the targeted minimal share for social housing in the municipality).[25] Neighbourhoods in this cluster almost entirely consist of newly built houses. Thus, for the future construction of such neighbourhoods, the local government could actively influence zoning plans, by, e.g., determining the proportion of social housing. Ideally,

24 We acknowledge that our model generally under-predicts the staying power for neighbourhoods when comparing predicted probabilities to those reported in the transition matrix in Figure 2 Most likely, some unobserved preferences such as location and housing wishes, which we cannot measure in our setting, also influence neighbourhood choice. Our simulations are thus to be seen as a "lower bound" to the expected staying power in a particular neighbourhood.

25 We set the ethnicity of this household to non-Western, as this is the cluster with the highest share of this ethnic group (31%). Our effects are robust with regard to the choice of ethnic group.

Figure 5
Simulation 2a: Response to an increase in social housing in the "new family houses" cluster on relocation choice of households from the "vulnerable" cluster



Note: The figure illustrates the probabilities of an average household from the "vulnerable" cluster to move to one of the nine neighbourhood clusters (blue bars), and how the same household responds to a change in neighbourhood characteristics for the "new family houses" cluster (i.e., an increase in social housing to 25%). Labels in the plot show the percentage change in estimated probabilities between the blue and grey bars.

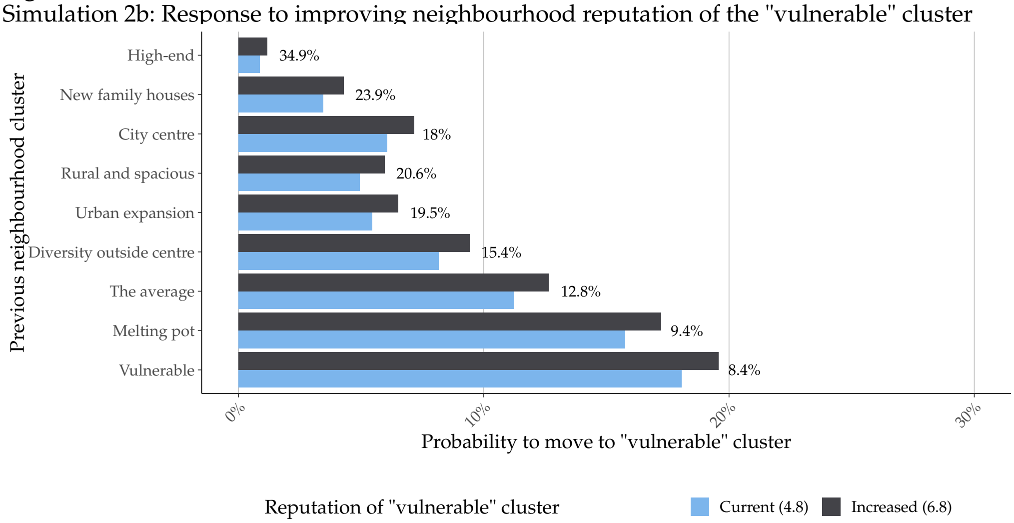
such policy would "break the spell" of households staying in problematic neighbourhoods, and allow some of the affected households to move elsewhere.

Figure 5 shows the results of our simulation. Specifically, for each of the nine neighbourhood clusters, we plot the choice probabilities for an average household originating from the "vulnerable" cluster for any of the available neighbourhoods (blue). We can now compare these choice probabilities to a situation in which the affordability of the "new family houses" cluster is increased through means of social housing. Indeed, our simulation shows that households respond strongly to the increased availability of social houses in that cluster, with choice probabilities increasing by 42% from 3.4% to 4.8%.

Simulation 2b: Improving neighbourhood reputation. Another tool for local governments to influence the moving behavior of their population is by improving neighbourhood reputation. Therefore, in this simulation, we assume that the municipality invested in measures to increase the neighbourhood reputation of the "vulnerable" cluster by, e.g., housing policies or improving neighbourhood amenities.

We consequently simulate how a rise of the neighbourhood reputation from 4.8 (its current value) to 6.8 (the average value in 's-Hertogenbosch) influenced the probability of moving to this cluster. If neighbourhood reputation mattered in household choices, then we would expect a strong response to this policy change. Figure 6 provides the results of this simulation. Note that, in comparison to earlier plots, this one illustrates the probability to move to the "vulnerable" cluster as a function of the "originating" neighbourhood. The results show that the probability of moving to the "vulnerable" cluster (with a better reputation) *increases* for households from *all* neighbourhood types. Interestingly, we observe the strongest increase for the neighbourhood clusters with

Figure 6
Simulation 2b: Response to improving neighbourhood reputation of the "vulnerable" cluster



Note: The figure illustrates the probabilities to move to the "vulnerable" cluster (reputation: 4.8, its current value) for an average household (blue bars), and for the same household moving to the "vulnerable" cluster with improved reputation (reputation: 6.8, the average value observed in the data; grey bars). Labels in the plot show the percentage change in estimated probabilities between the blue and grey bars.

the highest neighbourhood reputation. For example, the probability that households from the "new family houses" cluster and "high end" cluster relocate to the "vulnerable" cluster increases by 24% and 35%, respectively. This means that neighbourhoods which are less attractive for richer households can benefit from interventions which improve their liveability and thus diversify in terms of the neighbourhood population.

6. Discussion

This study investigated why households choose a particular neighbourhood type out of a set of alternative neighbourhood types. Below, we summarise our results and discuss the implications thereof, along the three key research questions raised in the introduction.

6.1 How can heterogeneous neighbourhoods be clustered?

As a first research question, we asked how potentially heterogeneous neighbourhoods can be clustered into more homogeneous neighbourhood types. To this extent, we first conducted an extensive literature review to identify variables of interest that describe neighbourhoods and neighbourhood choice. We then applied k-means clustering to the data. Based on the gap statistic, we obtained nine neighbourhood clusters. Similar to other studies, we observed some form of neighbourhood segregation (Hedman, Van Ham, and Manley 2011; Hedman and van Ham 2012). More disadvantaged neighbourhoods are characterised by a combination of low household income, low average dwelling values, high concentration of ethnic minorities and a very high share of social housing. Wealthier neighbourhoods have opposite characteristics and generally also experience higher staying power.

To the best of our knowledge, our study is the first study on neighbourhood choice that employs an unsupervised machine learning algorithm to grouping neighbourhoods. The existing literature has predominantly relied on *preconceived* neighbourhood classifications (e.g., based on a few socioeconomic indicators), or merely on administrative (and geographically-defined) borders (Clark, Deurloo, and Dieleman 2006).

Importantly, the emerging clusters did not overlap with the administrative districts defined by the municipality of ’s-Hertogenbosch, which to this date have been used predominantly as a way to group neighbourhoods and set policies. To the contrary, as we learnt from Figure 1, neighbourhoods of the same type are *scattered* across the entire municipality and not necessarily located close to each other. Thus, local governments can address policies specific to a particular neighbourhood type across the whole city, instead of relying on the predefined administrative division of districts which are more heterogeneous entities.

6.2 Which factors affect neighbourhood choice?

Second, we analysed which factors explained households’ neighbourhood choice, based on the interaction of household and neighbourhood characteristics. The results of the conditional logit model suggested that neighbourhood selection is not a structured process in the municipality of ’s-Hertogenbosch. While household income plays an important role, contrary to previous studies (Hedman, Van Ham, and Manley 2011; Schelling 1971; Boschman and Van Ham 2015), we did not find a dominant effect of household ethnicity, especially for households with a non-Western migration background. Also, in terms of other household characteristics (such as household composition and age), we did not notice a distinct pattern of neighbourhood selection.

This outcome is surprising, and potentially the result of previous housing policies of the municipality of ’s-Hertogenbosch, in which diversification of vulnerable neighbourhoods has received high priority. For example, new neighbourhoods ("new family houses" cluster) have a relatively high proportion of social housing, compared to older neighbourhoods.

We contribute to the academic literature by extending existing models on neighbourhood choice, which to this date have only focused on income and population composition in terms of ethnicity and household type (Hedman, Van Ham, and Manley 2011; Van Ham, Boschman, and Vogel 2018). We included covariates describing the physical environment, the structure of the housing market and neighbourhood reputation. Importantly, the new characteristics added to the model are all neighbourhood characteristics that local governments can *actively influence through their policy measures*, such as zoning plans.

6.3 By how much do factors affect neighbourhood choice?

As a third question, we asked by *how much changes in household and neighbourhood characteristics* influence the choice for a particular neighbourhood type. While multinomial choice models are well suited for explaining choice, the coefficient estimates tend to be hard to interpret - especially in our case, which has used multiple interaction terms at the neighbourhood and household level. Therefore, we used our model to simulate the effect of changing household and neighbourhood characteristics on the choice probabilities of affected neighbourhoods. We zoomed in on a few selected scenarios that are of interest to researchers in the domain of neighbourhood choice, and to decision-makers at the municipality that seek to make their city an attractive place for households to live.

On the one hand, our simulations revealed that as the economic situation of households deteriorates, the most vulnerable and least liveable neighbourhoods risk even further income segregation. On the other hand, the simulations showed that the addition of social housing in wealthier neighbourhoods would increase their attractiveness for households that previously lived in more vulnerable neighbourhoods. Those outcomes support the findings of Clark, Deurloo, and Dieleman (2006) and Hedman, Van Ham, and Manley (2011) that the staying power of neighbourhoods is not only based on preferences, but also on the constraints that households face. In other words, households with limited financial resources can not afford to relocate to other neighbourhoods, other than their own cluster, because housing is not affordable elsewhere (Clark, Deurloo, and Dieleman 2006; Hedman, Van Ham, and Manley 2011).

On the basis of our results, we recommend local governments to devote resources into the diversification of the housing market structure of both the more vulnerable and wealthier neighbourhoods as a potential remedy to prevent further segregation of neighbourhoods. For example, local governments could invest in less reputable neighbourhoods by making them more attractive for households with higher incomes. Additionally, the proportion of affordable housing could be enlarged in wealthier neighbourhoods to offer opportunities for households with less financial resources to relocate to those neighbourhoods. The outcomes of our simulations suggest that households potentially relocate to other neighbourhoods as a response to those measures. Above interventions by local governments could be possible solutions allowing those households to also transfer to different types of neighbourhoods.

The addition of a simulation study is unique to the housing literature. For one, we can shed light on the magnitude of the effects. However, our simulations can also be used by policymakers to forecast the impact of new policy measures on the moving behaviour of existing (and potentially new) inhabitants of their municipality.

6.4 Opportunities for data-driven research

We have used the multinomial choice model to simulate the behaviour of households as a response to changes to household or neighbourhood characteristics. We would like to acknowledge that the multinomial choice model traditionally is used for hypothesis testing (i.e., confirmatory statistics), and may show weaker performance in terms of obtaining accurate predictions. However, decision-makers may not only be interested in which factors affect choice but also require the best possible predictions given a set of (new) policy measures.

In the following, we briefly give an overview of a potential set up of such a predictive approach. In the context of a learning-oriented methods, we consider the prediction of neighbourhood choices as a multiclass classification problem which treats the different neighbourhood types as target labels, and the neighbourhood and household characteristics as features. As we are dealing with an unbalanced data set, i.e. we observe many residential moves for some neighbourhood clusters but only a few for others, we recommend to balance the target data before model estimation. Also, we would ideally split our sample in a training set (which we would use for learning), and a validation set (allowing us to measure the accuracy of the model on unseen data). A general risk of applying learning algorithms is overfitting, i.e. that the model does not generalise well on unseen data. Various techniques of regularisation, such as early stopping, exist which are beyond the scope of this section.

Potential methods suitable to predicting location choice are decision trees and neural networks[26]. Decision trees are very transparent learning techniques which predict the class of a new observation based on decision rules learnt from the data. The advantage of this approach is that it allows us to *explain and interpret* the outcomes, as we can reiterate an example through the tree. In other words, we would be able to understand which household and neighbourhood features triggered the prediction towards one particular neighbourhood label. The drawback of this method is that it does not show how choice probabilities of alternative neighbourhoods respond to a change in neighbourhood or household characteristics.

A feedforward neural network adds hidden layers between the input features (i.e., neighbourhood and household characteristics) and the target (i.e., the chosen neighbourhood), making it a powerful learning technique often performing well in terms of prediction accuracy. It uses a softmax activation function in the output layer to determine the output. The advantage of a neural network is the automatic feature selection which is achieved through the learning approach. This merit of neural networks would allow us to extend further the list of variables used in our conditional logit model, e.g., by attributes of the previous housing of a household. Because of the complexity of neural networks, however, the results are more difficult to interpret and might not be transparent enough from the perspective of local governments, that need to justify and be open about their decision-making processes.

6.5 Limitations and future research

Even within the boundaries of the literature on neighbourhood choice, this study faces limitations concerning data quality and scope. We had to approximate household income based on information on age categories or neighbourhood level. As a result, the diversity of income levels within a neighbourhood cluster may not be completely reflected. Additionally, we have not modelled housing supply, but acknowledge this certainly is an essential driver of neighbourhood choice. While it is crucial for local governments to understand which type of households are leaving the municipality, we only could include moves into and within the municipality. Analysing households relocating elsewhere would allow decision-makers to respond to the particular demand of those households in terms of housing needs and wishes and implement policies which prevent them from leaving.

A critical aspect of our study was the clustering procedure used to group neighbourhoods into more homogeneous neighbourhood types by using the k-means clustering approach. While k-means clustering is a powerful unsupervised learning technique, it relies on manually setting k, i.e. the number of clusters. While the gap-statistic used in this study helps to define k, ideally we would like the algorithm to learn the number of groups. An alternative approach, which incorporates determining the number of clusters, is hierarchical clustering. This algorithm initially treats all neighbourhoods as single clusters. Based on the distance between those clusters, the two most similar neighbourhoods are merged into one cluster. This process is repeated until all initial clusters group together in one cluster. A dendrogram subsequently illustrates the hierarchy between the iteratively grouped neighbourhoods and allows to determine the preferred number of clusters.

26 Several other classifiers are suitable for multiclass prediction problems, e.g., k-nearest neighbour, naive Bayes and support vector machines. It is beyond the scope of this discussion to introduce all of those.

For the conditional logit model estimated in this study, we used a literature-driven approach to select variables for model inclusion. Since the conditional logit model limits the number of potential covariates to add, we needed to reduce the number variables to what we believed were the most relevant ones. Future work may use a more sophisticated model selection process to explore the wealth of different variable combinations, and eventually, choose the one with the best model fit.

To investigate the drivers influence neighbourhood choice, we applied a purely explanatory approach. In particular, we used the well-known multinominal logit model, which is the preferred method across many disciplines when studying choice with more than two alternatives (Greene 2012). Yet, next to understanding the drivers influencing neighbourhood choices, municipalities also need to forecast residential location choices as a function of policy changes. While we used our model to some extent for predictive purposes by simulating how potential policy changes affect neighbourhood choices, the forecasting power of our model is limited (pseudo-R^2 of 12.5%). We believe that a more exploratory and data science-oriented approach, such as decision trees or neural networks mentioned earlier, may offer ample opportunities to boost model fit, and obtain more accurate forecasts for decision making.

7. Conclusion

Against the backdrop of housing shortages, local governments need effective housing policies to stimulate the growth of housing supply. To understand how the implementation of those policies affects the housing demand of consumers, this study analysed the residential mobility of households in the Dutch municipality of 's-Hertogenbosch between the years 2014 to 2018. Conceptually, residential mobility can be explained by the reason *why* households move, and the decision of *where* to move to. We focused on the latter by modelling the residential location choice of households. As our first research question, we asked how heterogeneous neighbourhoods can be grouped into more homogeneous neighbourhood types. Second, we specified a conditional logit model to ask which neighbourhood and household characteristics drive neighbourhood choice at the household level. Last, we used the model to simulate how changes in the socio-economic situation of households and potential government measures influence neighbourhood selection.

We identified nine neighbourhood clusters which spread across the border of the municipality districts. We found that most clusters have a high staying power, which means that households move within the same neighbourhood clusters frequently. The analysis of neighbourhood choices based on neighbourhood and household characteristics emphasised that household income plays an important role. In opposition to previous studies, we did not find a strong effect of household ethnicity. We also did not observe a distinct pattern of neighbourhood choice in terms of other household characteristics (such as household composition and age). Simulations showed that the most vulnerable neighbourhoods risk further segregation if the economic situation of households deteriorated. Using different policy measures, we illustrated how our model could be used to understand how measures to increase neighbourhood diversification affect neighbourhood selection.

From a policy-maker perspective, it is crucial to both explain and predict neighbourhood choices. Since the model applied in this study enables us to combine those aspects to some degree, despite the limitations mentioned above, we believe to have built a rich basis for both social scientists and data scientists to conduct follow-up research.

References

Boschman, Sanne and Maarten Van Ham. 2015. Neighbourhood selection of non-Western ethnic minorities: Testing the own-group effects hypothesis using a conditional logit model. *Environment and Planning A*, 47(5):1155–1174.

Brown, Lawrence A and Eric G Moore. 1970. The intra-urban migration process: a perspective. *Geografiska Annaler: Series B, Human Geography*, 52(1):1–13.

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Multinomial Models*. Cambridge University Press.

Chhetri, Prem, Robert J Stimson, and John Western. 2006. Modelling the factors of neighbourhood attractiveness reflected in residential location decision choices. *Studies in Regional Science*, 36(2):393–417.

Clark, William, Marinus Deurloo, and Frans Dieleman. 2006. Residential mobility and neighbourhood outcomes. *Housing Studies*, 21(3):323–342.

Clark, William AV and Valerie Ledwith. 2006. Mobility, housing stress, and neighborhood contexts: evidence from Los Angeles. *Environment and Planning A*, 38(6):1077–1093.

Coulter, Rory and Jacqueline Scott. 2015. What motivates residential mobility? Re-examining self-reported reasons for desiring and making residential moves. *Population, Space and Place*, 21(4):354–371.

Dieleman, Frans M. 2001. Modelling residential mobility; a review of recent trends in research. *Journal of Housing and the Built Environment*, 16(3-4):249–265.

Greene, W.H. 2012. *Econometric Analysis*. Pearson International Edition. Pearson Education, Limited.

de Groot, C and N Vrieselaar. 2019. Housing shortage continues to drive up house prices. https://economics.rabobank.com/publications/2019/november/housing-shortage-continues-to-drive-up-house-prices/. Accessed: 2020-06-03.

Hedman, Lina and Maarten van Ham. 2012. Understanding neighbourhood effects: Selection bias and residential mobility. In *Neighbourhood Effects Research: New Perspectives*. Springer, pages 79–99.

Hedman, Lina, Maarten Van Ham, and David Manley. 2011. Neighbourhood choice and neighbourhood reproduction. *Environment and Planning A*, 43(6):1381–1399.

Hoffman, Saul D and Greg J Duncan. 1988. Multinomial and conditional logit discrete-choice models in demography. *Demography*, 25(3):415–427.

Ioannides, Yannis M, Jeffrey E Zabel, et al. 2008. Interactions, neighborhood selection and housing demand. *Journal of urban economics*, 63(1):229.

Kleinepier, T, K Gopal, D Omtzigt, and G van Leeuwen. 2019. Primos 2019: Prognose van bevolking, huishoudens en woningbehoefte 2018-2050. Technical Report r2019-0055TK, ABF Research.

Lee, Monghyeon and Yongwan Chun. 2016. Residential relocation in a Metropolitan Area: A case study of the Seoul Metropolitan Area, South Korea. In *Spatial Econometric Interaction Modelling*. Springer, pages 441–463.

Mann, Richard P, Viktoria Spaiser, Lina Hedman, and David JT Sumpter. 2018. Choice modelling with Gaussian processes in the social sciences: A case study of neighbourhood choice in Stockholm. *PloS one*, 13(11).

McFadden, Daniel. 1974. The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328.

Morris, Timothy. 2017. Examining the influence of major life events as drivers of residential mobility and neighbourhood transitions. *Demographic Research*, 36:1015–1038.

Permentier, Matthieu, Maarten Van Ham, and Gideon Bolt. 2009. Neighbourhood reputation and the intention to leave the neighbourhood. *Environment and Planning A*, 41(9):2162–2180.

Sánchez, Aida Caldera and Dan Andrews. 2011. Residential mobility and public policy in OECD countries. *OECD Journal: Economic Studies*, 2011(1):1–22.

Schelling, Thomas C. 1971. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Thomas, Michael, John Stillwell, and Myles Gould. 2016. Modelling mover/stayer characteristics across the life course using a large commercial sample. *Population, Space and Place*, 22(6):584–598.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

Van Ham, Maarten, Sanne Boschman, and Matt Vogel. 2018. Incorporating neighborhood choice in a model of neighborhood effects on income. *Demography*, 55(3):1069–1090.
Van Ham, Maarten and William AV Clark. 2009. Neighbourhood mobility in context: household moves and changing neighbourhoods in the Netherlands. *Environment and planning A*, 41(6):1442–1459.
Van Ham, Maarten and Peteke Feijten. 2008. Who wants to leave the neighbourhood? The effect of being different from the neighbourhood population on wishes to move. *Environment and Planning A*, 40(5):1151–1170.
Van der Vlist, Arno J, Cees Gorter, Peter Nijkamp, and Piet Rietveld. 2002. Residential mobility and local housing-market differences. *Environment and Planning A*, 34(7):1147–1164.
WRR. 2005. *Vertrouwen in de buurt*, volume 72. Amsterdam University Press.

Appendix

A. Cluster description

The "vulnerable" cluster receives the lowest score in terms of the neighbourhood reputation (4.8). It witnesses by far the highest share of non-western minorities (31%) and also the lowest households income (21,900€). 75% of the housing stock is occupied by social housing. The "melting pot" cluster stands out by its diversity in terms of income levels and ethnic composition as well as structure of the housing market. Being close to the city centre, this cluster attracts both non-western (17%) as well as western minorities (10%) to live there. While 63% of the housing market is occupied by social housing, 12% is private rental, which is slightly above average. For "the average" cluster we find many of its characteristics to describe the average neighbourhood of ′s-Hertogenbosch. With a mean household income of 28,000€, almost 11% households with a non-western migration background, an almost even split between singles, couples and family with children and a neighbourhood reputation of score 7 it represents very much the average neighbourhood of ′s-Hertogenbosch. Also in terms of housing density and distance from the central train station and access to highway access lane, its situated in the middle.

While the "diversity outside the city centre" cluster is close to the city centre (2.4km) and has similar services available (151 versus 160 restaurants within 3km), it offers bigger and cheaper houses. This cluster combines diverse neighbourhoods which lie just outside the city centre. As houses are larger and the housing density is lower, this cluster is also more attractive to families with children than the more central clusters (26%). The "urban expansion" cluster consist of neighbourhoods which were former village centres or were integrated through urban expansion in the 1970s and 1980s. This cluster mainly consist of dwellings built before 2000, which are mostly owner-occupied homes (76%). Both mean household income (34.800 €) and average housing size (145 m^2) are above average. The "rural and spacious" cluster, which is located far from the city centre (12 km), has by far the lowest housing density (150 addresses per km^2). Houses are expensive (410,000€) and large (187 m^2). The share of ethnic minorities is low (3.8%) and the housing market is dominated by home-ownership (84%).

The "city centre" is dominated by single households (63%) and a low share of home-ownership (36%). Furthermore, 12% of the inhabitants belong to Western minorities. Naturally, this cluster is characterised by its closeness to the central train station (1.1 km) and the high number of restaurants in the vicinity (160). The "new family houses cluster" stands out by the high share of newly built houses (90%), which are relatively big (173 m^2) and which are mainly occupied by families with children (57%). Those neighbourhoods receive a very high score in terms of neighbourhood reputation (8.6). The "high-end cluster" is characterised by the highest mean property value (619,000€), the highest mean household income, a high concentration of native Dutch households (94%), and its excellent neighbourhood reputation (9). With only two neighbourhoods and 445 households living there, this is also the smallest cluster. The structure of the housing market (with 96% of home ownership and large and expensive dwellings) explain the low turnover rate, as these houses are only affordable for a small group of households.

Table 7
Mean characteristics of neighbourhood clusters (2018, N = 72 neighbourhoods)

	Vulnerable	Melting pot	The average	Diversity outside centre	Urban expan-sion	Rural and spacious	City centre	New family houses	High-end
Number of neighbourhoods	8	13	10	7	16	4	4	8	2
Total number of households	7,780	13,005	8,940	9,080	15,365	1,085	8,935	5,695	445
Mean number of households	972.5	1,000.4	894	1,297.1	960.3	271.2	2,233.8	711.9	222.5
Mean number of inhabitants	2,197.5	1,808.8	1,890.5	2,547.9	2,341.6	716.2	3,410	2,028.1	612.5
Turnover rate (in %)	5.8	12.7	7	8.7	6.8	5.9	9.8	8.6	4.1
Neighbourhood reputation*	4.8	5.3	6.6	6.6	7.7	8	8	8.6	9
Average dwelling values (x 1000, in €)*	173	189.5	215.4	249.6	308.7	409.5	273	320.8	619
Mean household income (x 1000, in €)*	21.9	24.1	28.2	30.4	34.8	37.1	33.4	35.9	48
Share of native Dutch (in %)	58.5	73.5	79.3	83	89.1	96.2	79.5	86.5	93.5
Share of non-western minorities (in %)*	31.4	16.8	10.2	8.6	3.9	0.8	8.2	6.1	1.5
Share of western minorities (in %)*	10.1	9.6	10.5	8.4	7	3	12.2	7.4	5
Share of singles or Other (in %)*	37.4	55.7	38.9	45.9	25.7	21	62.5	18.2	13
Share of couples (in %)	24.2	24.3	29.7	28	34.8	36.2	28.5	26.2	45
Share of families with children (in %)*	39	20.9	32.3	26.4	40	46	9.5	56.6	44.5
Mean housing size (in m^2)	116	97.3	110.1	103.3	144.9	186.5	94.8	173	264
Share of home-ownership (in %)	22.4	24	55.5	54.6	75.9	83.8	35.5	72.2	96
Share of private rental (in %)*	2.2	12.2	5.9	15.6	9.1	10.5	33.2	10.1	3.5
Share of social housing (in %)*	75.1	62.7	37.3	29	14.1	4.8	29.2	17.5	0
Share of new houses (in %)*	4	15.8	18.5	7.9	6.8	12.5	25.2	90.2	12.5
Housing density*	1,733.6	3,124.7	1,621.4	2,387.1	1,138.1	150	3,594.2	794.4	1,278.5
Distance to highway (in km)*	1.2	2.2	1.8	2	1.5	3.1	2.7	2	1.9
Distance to train station (in km)*	3.5	2.1	5.1	2.4	6.5	11.5	1.1	6.4	2.5
Restaurants within 3 km*	36.1	150.3	13.4	150.9	12.3	3.8	160.3	23	114.9

* Variables used for k-means clustering. Other variables added for further description of neighbourhood clusters.

B. Results from conditional logit model

Table 8: Results of conditional logit model on neighbourhood choice: Baseline model versus full model.

	Dependent variable:	
	Neighbourhood choice	
	Baseline model	Full model
Interactions with Average dwelling values		
Moves within *(Reference: Moves into)*		-2.616^{***}
		(0.153)
Household income		-0.399^{***}
		(0.093)
Non-western minority *(Reference: Native Dutch)*		-0.484
		(0.452)
Western minority *(Reference: Native Dutch)*		0.280
		(0.319)
Singles or other *(Reference: Couples)*		-1.108^{***}
		(0.230)
Families with children *(Reference: Couples)*		-0.386^{*}
		(0.209)
< 25 years *(Reference: 25 - 44 years)*		-0.412
		(0.544)
45 - 64 years *(Reference: 25 - 44 years)*		-0.195
		(0.226)
> 65 Years *(Reference: 25 - 44 years)*		-1.462^{***}
		(0.354)
Room-stress		-0.073
		(0.135)
Interactions with Mean neighbourhood income		
Moves within *(Reference: Moves into)*		2.027^{***}
		(0.218)
Household income	0.187^{***}	0.342^{**}
	(0.008)	(0.138)
Non-western minority *(Reference: Native Dutch)*		0.541
		(0.606)
Western minority *(Reference: Native Dutch)*		-1.185^{**}
		(0.502)
Singles or other *(Reference: Couples)*		0.809^{**}
		(0.317)
Families with children *(Reference: Couples)*		0.608^{**}
		(0.297)
< 25 years *(Reference: 25 - 44 years)*		-0.765
		(0.714)
45 - 64 years *(Reference: 25 - 44 years)*		-0.437
		(0.319)

> 65 Years *(Reference: 25 - 44 years)*		0.456
		(0.485)
Room-stress		−0.009
		(0.192)
Interactions with Share of non-western minorities		
Moves within *(Reference: Moves into)*		0.062
		(0.045)
Household income		0.041
		(0.029)
Non-western minority *(Reference: Native Dutch)*	0.306***	0.073
	(0.025)	(0.093)
Western minority *(Reference: Native Dutch)*	0.039	0.222**
	(0.029)	(0.093)
Singles or other *(Reference: Couples)*		−0.057
		(0.056)
Families with children *(Reference: Couples)*		−0.071
		(0.059)
< 25 years *(Reference: 25 - 44 years)*		0.019
		(0.115)
45 - 64 years *(Reference: 25 - 44 years)*		−0.067
		(0.061)
> 65 Years *(Reference: 25 - 44 years)*		−0.134*
		(0.077)
Room-stress		−0.064*
		(0.034)
Interactions with Share of western minorities		
Moves within *(Reference: Moves into)*		−1.044***
		(0.080)
Household income		−0.224***
		(0.049)
Non-western minority *(Reference: Native Dutch)*	0.462***	−0.038
	(0.055)	(0.177)
Western minority *(Reference: Native Dutch)*	0.459***	−0.181
	(0.047)	(0.171)
Singles or other *(Reference: Couples)*		0.021
		(0.099)
Families with children *(Reference: Couples)*		−0.270**
		(0.105)
< 25 years *(Reference: 25 - 44 years)*		−0.009
		(0.215)
45 - 64 years *(Reference: 25 - 44 years)*		0.092
		(0.105)
> 65 Years *(Reference: 25 - 44 years)*		−0.350**
		(0.139)
Room-stress		0.048
		(0.063)
Interactions with Share of singles or other		

Moves within *(Reference: Moves into)*		-0.345^{***}
		(0.101)
Household income		-0.063
		(0.060)
Non-western minority *(Reference: Native Dutch)*		0.073
		(0.283)
Western minority *(Reference: Native Dutch)*		-0.771^{**}
		(0.310)
Singles or other *(Reference: Couples)*	0.249^{***}	-0.178
	(0.036)	(0.136)
Families with children *(Reference: Couples)*	0.155^{***}	0.350^{***}
	(0.034)	(0.135)
< 25 years *(Reference: 25 - 44 years)*		0.560^{*}
		(0.297)
45 - 64 years *(Reference: 25 - 44 years)*		0.296^{**}
		(0.138)
> 65 Years *(Reference: 25 - 44 years)*		-0.313
		(0.257)
Room-stress		-0.152^{*}
		(0.092)
Interactions with Share of families with children		
Moves within *(Reference: Moves into)*		-1.150^{***}
		(0.128)
Household income		-0.223^{***}
		(0.079)
Non-western minority *(Reference: Native Dutch)*		-0.020
		(0.358)
Western minority *(Reference: Native Dutch)*		-1.035^{**}
		(0.414)
Singles or other *(Reference: Couples)*	-0.405^{***}	-0.319^{*}
	(0.036)	(0.164)
Families with children *(Reference: Couples)*	0.221^{***}	0.119
	(0.033)	(0.160)
< 25 years *(Reference: 25 - 44 years)*		0.491
		(0.351)
45 - 64 years *(Reference: 25 - 44 years)*		0.583^{***}
		(0.167)
> 65 Years *(Reference: 25 - 44 years)*		-0.390
		(0.320)
Room-stress		0.002
		(0.114)
Interactions with Share of social housing		
Moves within *(Reference: Moves into)*		0.292^{***}
		(0.088)
Household income		-0.014
		(0.054)

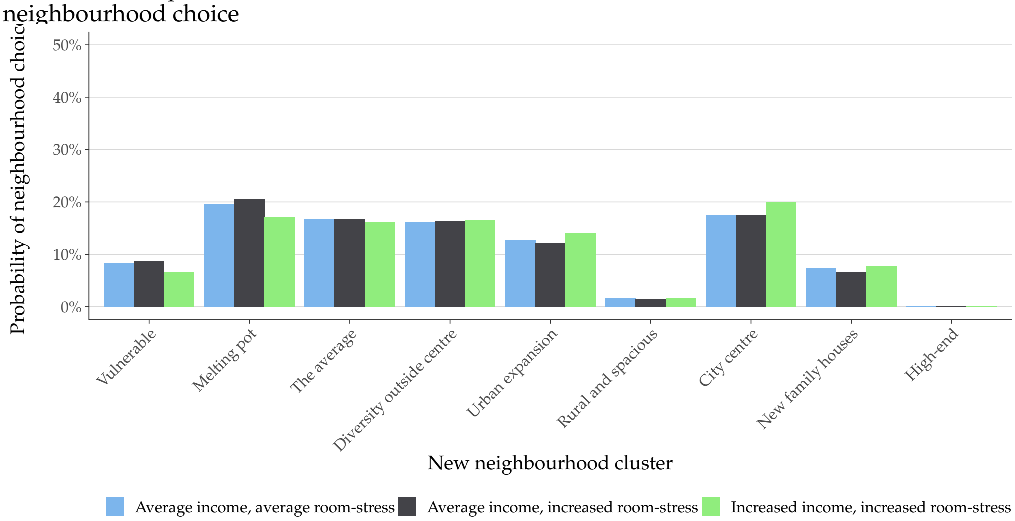
Non-western minority *(Reference: Native Dutch)*	0.712^{***}
	(0.213)
Western minority *(Reference: Native Dutch)*	-0.420^{**}
	(0.184)
Singles or other *(Reference: Couples)*	0.183
	(0.117)
Families with children *(Reference: Couples)*	-0.040
	(0.120)
< 25 years *(Reference: 25 - 44 years)*	-0.031
	(0.257)
45 - 64 years *(Reference: 25 - 44 years)*	-0.167
	(0.122)
> 65 Years *(Reference: 25 - 44 years)*	0.214
	(0.167)
Room-stress	0.162^{**}
	(0.074)
Interactions with Share of new houses	
Moves within *(Reference: Moves into)*	0.327^{***}
	(0.052)
Household income	0.025
	(0.032)
Non-western minority *(Reference: Native Dutch)*	0.135
	(0.123)
Western minority *(Reference: Native Dutch)*	0.292^{**}
	(0.132)
Singles or other *(Reference: Couples)*	0.009
	(0.066)
Families with children *(Reference: Couples)*	0.197^{***}
	(0.068)
< 25 years *(Reference: 25 - 44 years)*	-0.421^{***}
	(0.142)
45 - 64 years *(Reference: 25 - 44 years)*	-0.241^{***}
	(0.069)
> 65 Years *(Reference: 25 - 44 years)*	-0.236^{**}
	(0.105)
Room-stress	-0.103^{**}
	(0.040)
Interactions with Distance to highway	
Moves within *(Reference: Moves into)*	-0.437^{***}
	(0.036)
Household income	-0.085^{***}
	(0.022)
Non-western minority *(Reference: Native Dutch)*	-0.051
	(0.097)
Western minority *(Reference: Native Dutch)*	-0.168^{*}
	(0.091)
Singles or other *(Reference: Couples)*	0.171^{***}

	(0.047)
Families with children *(Reference: Couples)*	-0.401^{***}
	(0.050)
< 25 years *(Reference: 25 - 44 years)*	0.194^{*}
	(0.103)
45 - 64 years *(Reference: 25 - 44 years)*	0.137^{***}
	(0.050)
> 65 Years *(Reference: 25 - 44 years)*	-0.130^{*}
	(0.069)
Room-stress	0.079^{**}
	(0.031)
Interactions with Distance to train station	
Moves within *(Reference: Moves into)*	0.214^{***}
	(0.047)
Household income	0.005
	(0.029)
Non-western minority *(Reference: Native Dutch)*	−0.003
	(0.122)
Western minority *(Reference: Native Dutch)*	-0.224^{**}
	(0.092)
Singles or other *(Reference: Couples)*	0.080
	(0.064)
Families with children *(Reference: Couples)*	0.397^{***}
	(0.063)
< 25 years *(Reference: 25 - 44 years)*	−0.217
	(0.149)
45 - 64 years *(Reference: 25 - 44 years)*	-0.196^{***}
	(0.066)
> 65 Years *(Reference: 25 - 44 years)*	0.280^{***}
	(0.088)
Room-stress	-0.104^{***}
	(0.039)
Interactions with Neighbourhood reputation	
Moves within *(Reference: Moves into)*	0.120
	(0.129)
Household income	0.334^{***}
	(0.079)
Non-western minority *(Reference: Native Dutch)*	0.120
	(0.275)
Western minority *(Reference: Native Dutch)*	0.281
	(0.251)
Singles or other *(Reference: Couples)*	−0.240
	(0.166)
Families with children *(Reference: Couples)*	−0.274
	(0.180)
< 25 years *(Reference: 25 - 44 years)*	0.980^{***}
	(0.354)

45 - 64 years *(Reference: 25 - 44 years)*		0.049
		(0.180)
> 65 Years *(Reference: 25 - 44 years)*		0.328
		(0.221)
Room-stress		0.028
		(0.095)
Observations	18,318	18,318
Log Likelihood	-38,577.320	-34,937.830
Pseudo-R^2	0.034	0.125
Note:		*p<0.1; **p<0.05; ***p<0.01

C. Simulations

Figure 7
Simulation: Response to a 25%-decrease in room-stress and 10% income increase on neighbourhood choice



Note: The figure illustrates the probabilities of moving to one of the nine neighbourhood clusters for an average household with average room-stress (blue bars), the same average household with 25% more room-stress (grey bars), and the same average household with 10% higher income and 25% more room-stress (green bars).